A project report on

WILDFIRE DETECTION USING VISION TRANSFORMERS

By Aarya Mahajan Pabha

ABSTRACT

Early wildfire detection plays a crucial role in minimizing property damage and saving lives. This thesis explores the potential of Vision Transformers (ViTs) for wildfire image classification and subclassification. We propose a novel cascading ViT-based system that leverages the strengths of ViTs in feature extraction and classification.

The system employs a binary ViT classifier to distinguish fire from no-fire scenarios. Subsequently, dedicated submodels handle fire and no-fire subclass classification. Our evaluations demonstrate the effectiveness of the binary ViT model, achieving a test accuracy of 97.8%. While experiments explored various ViT architectures for no-fire subclassification, none surpassed the binary classification accuracy (92.43%). This suggests a potential benefit for using the ViT model approach over other transformers.

The cascading architecture with dedicated submodels achieved a promising overall test accuracy of 88.54% with detailed class-wise performance presented in a confusion matrix. Furthermore, the proposed Binary ViT model (10 epochs) outperformed a published method ("Dual-Dataset Deep Learning for Improved Forest Fire Detection") in terms of accuracy, precision, and specificity.

This work contributes to wildfire detection by demonstrating the effectiveness of ViTs and introducing a cascading system for both fire/no-fire classification and subclassification.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

- ROC-AUC (Receiver Operating Characteristic - Area Under the Curve)
- MCC (Matthews Correlation Coefficient)
- FNR (False Negative Rate)
- F1-Score
- FN (False Negative)
- TP (True Positive)
- TN (True Negative)
- FP (False Positive)
- STD (Standard Deviation)
- CNN (Convolutional Neural Network)
- ViT (Vision Transformer)

KEYWORDS

Table 1.1 Keywords

Keyword	Definition
Binary Model	A machine learning model trained to classify data into two categories (e.g., fire vs. no-fire).
Cascading ViT System	A deep learning system that utilizes multiple ViT models in a sequential manner for image classification and subclassification.
Confusion Matrix	A table that visualizes the performance of a classification model on a test dataset, showcasing the number of correctly and incorrectly classified instances for each class.
Early Wildfire Detection	The process of identifying wildfires at an early stage to minimize damage and enable timely response.
False Negative Rate (FNR)	The proportion of actual positive cases (e.g., fire images) that are incorrectly classified as negative (e.g., no-fire).
F1-Score	A harmonic mean of precision and recall, providing a balanced view of model performance.
Fire Subclass	A specific category within the fire class (e.g., Smoke_from_fires, Both_smoke_and_fire).
Hyperparameter	A setting that controls the training process of a machine learning model but is not directly learned from the data.
Matthews Correlation Coefficient (MCC)	A metric that considers true positives, true negatives, false positives, and false negatives to assess the overall performance of a binary classification model.
No-Fire Subclass	A specific category within the no-fire class (e.g., Forested_areas_without_confounding_elements, Fire_confounding_elements, Smoke_confounding_elements).

Precision	The proportion of predicted positive cases that are truly positive.
Recall	The proportion of actual positive cases that are correctly identified as positive.
Receiver Operating Characteristic (ROC) Curve	A graphical representation that illustrates the performance of a binary classification model at all classification thresholds.
Specificity	The proportion of actual negative cases that are correctly identified as negative.
Test Accuracy	The proportion of correctly classified images in the test dataset used to evaluate model performance.
Vision Transformer (ViT)	A deep learning model architecture designed for image classification tasks.

Chapter 1

Introduction

1.1 WILDFIRE DETECTION

The escalating intensity and frequency of wildfires pose a dire threat to global ecosystems, jeopardizing biodiversity, disrupting natural resource cycles, and causing significant economic and social losses. Early detection is paramount for effective mitigation and response, yet traditional methods like human observation and sensor networks suffer from limitations in range, accuracy, and real-time response capabilities. This thesis proposes that deep learning models, particularly Vision Transformers (ViTs), represent a revolutionary approach to wildfire detection. ViTs' exceptional ability to analyse vast amounts of visual data with unparalleled precision makes them ideally suited for real-time identification of wildfires based on smoke patterns, thermal anomalies, and other critical visual indicators. This research aims to develop and rigorously evaluate a novel ViT architecture specifically tailored for wildfire detection. This tailored architecture will be trained on a meticulously curated and augmented dataset of wildfire imagery, incorporating diverse environmental conditions and fire types. The success of this research has the potential to revolutionize wildfire management by enabling:

- **Real-time, wide-area detection:** ViT-based systems can continuously monitor vast landscapes, significantly reducing the reliance on human patrol limitations.
- **Enhanced accuracy and reduced false positives:** The precise analysis offered by ViTs can distinguish wildfires from other phenomena, leading to faster and more targeted responses.
- **Improved response times:** Early detection translates to faster deployment of firefighting resources, minimizing property damage and loss of life.

- **Proactive mitigation strategies:** Early detection allows for earlier intervention efforts, such as controlled burns, to manage fuel loads and potentially prevent catastrophic wildfires.

By achieving these advancements, this research can contribute significantly to the global effort towards mitigating the devastating consequences of wildfires, ultimately leading to a more resilient and sustainable future for our planet.

1.2 BACKGROUND

Wildfires are a growing menace across the globe, their frequency and intensity fueled by climate change, human activity, and natural phenomena like lightning strikes. Early detection is paramount in mitigating the devastating consequences of wildfires, which include property damage exceeding billions of dollars annually, loss of life, and ecological destruction. This study delves into the potential of deep learning models, particularly Vision Transformers (ViTs), to revolutionize wildfire detection by leveraging advancements in computer-vision.

1.3 CONTEXT

Traditional wildfire detection methods primarily rely on human observation from lookout towers or aerial patrols, and sensor networks strategically placed in high-risk areas. While these methods have served their purpose, they come with inherent limitations. Human observation is restricted by range and susceptible to fatigue or error. Sensor networks, often expensive to set up and maintain, might have limited coverage or require complex infrastructure, especially in remote-areas.

The emergence of deep learning has brought significant advancements in image recognition tasks. Vision Transformers (ViTs) represent a particularly promising approach. Unlike traditional Convolutional Neural Networks (CNNs), ViTs excel at analyzing long-range dependencies within images, making them ideal for tasks that require identifying subtle patterns indicative of wildfires, such as smoke plumes, thermal anomalies, and specific vegetation changes.

1.4 SIGNIFICANCE OF THE STUDY

This research proposes a novel Vision Transformer architecture specifically designed for accurate and real-time wildfire detection in imagery. The study goes beyond just detection by incorporating Explainability techniques. By making the model's decision-making process more transparent and interpretable, researchers aim to build trust with human operators who will ultimately rely on the system's recommendations.

The successful development of a reliable ViT-based wildfire detection system could have far-reaching positive impacts:

Enhanced Early Detection: Faster detection translates to faster response times by firefighters, potentially minimizing property damage, loss of life, and ecological devastation.

Improved Resource Allocation: More accurate information about wildfire location and spread allows for efficient allocation of firefighting resources, optimizing their impact.

Proactive Mitigation Strategies: Early detection opens doors for proactive measures like controlled burns, which can help manage fuel loads and potentially prevent catastrophic wildfires.

Wider Coverage and Reduced Costs: ViT-based systems can continuously monitor vast landscapes, potentially replacing some human patrols and reducing the reliance on expensive, fixed sensor networks.

1.5 PROBLEM STATEMENT

Driven by climate change, human activity, and increasingly volatile environmental conditions, wildfires are escalating in frequency, intensity, and unpredictability. These infernos wreak havoc on ecosystems, endanger human lives and infrastructure, and leave lasting scars on communities. Early detection is crucial for effective response and mitigation efforts, yet current methods, often reliant on human observation or sensor networks, are demonstrably slow and unreliable. This critical gap necessitates a radical paradigm shift in our ability to detect wildfires. This research proposes a novel approach leveraging cutting-edge Vision Transformer (ViT) technology for real-time wildfire detection in imagery. By harnessing the exceptional ability of ViTs to analyze vast amounts of visual data with high precision, this research aims to develop a more accurate and timely wildfire detection system. The successful development of such a system has the potential to save lives, protect our planet, and bolster our resilience in the face of the growing wildfire crisis.

Chapter 2

Literature Review

This research [1] explores deep learning for wildfire detection and segmentation using UAV imagery. They investigate Ensemble CNNs, a common deep learning technique, and emerging Vision Transformers for segmentation tasks. The study utilizes the publicly available FLAME dataset for training and evaluation, facilitating comparisons and result replicability. F1-score, accuracy, and inference time are used for performance assessment. While inference time is crucial for real-time applications, details are lacking.

The authors highlight the need for future research focused on optimizing models for real-time detection and segmentation. This aligns with the critical need for faster and more accurate wildfire detection systems. This paper contributes to the growing body of research on deep learning for wildfire detection, showcasing its potential while emphasizing the importance of real-time optimization.

This paper [2] emphasizes the critical role of early fire detection in power plant safety and proposes a novel approach utilizing deep learning. Their research focuses on an automated fire early warning system built using a Vision Transformer (ViT) model for image recognition. The study demonstrates the effectiveness of the ViT-based model through comparative analysis, highlighting its potential to prevent fire-related economic losses and casualties in power plants.

However, the study presents some limitations that warrant further exploration. First, the dataset used for training the ViT model relies on Google Photos, which raises concerns about generalizability. A more standardized and publicly available dataset would be preferable for robust model development and replicability. Second, the paper lacks a thorough discussion of the practical challenges in implementing this system in real-world power plant environments. Considerations such as computational resource requirements for real-time processing and scalability for large facilities need to be

addressed. Additionally, the potential impact of environmental factors like changing light conditions or the presence of non-fire heat sources on the model's accuracy and reliability requires further investigation to ensure its effectiveness in diverse operational scenarios.

This paper [3] investigates the application of deep learning, specifically deep semantic segmentation, for wildfire detection using multi-sensor satellite imagery. The research explores the potential of this approach in both clear and cloudy conditions, aiming to overcome limitations associated with traditional methods. The study evaluates the performance of various models in identifying fire-affected pixels, employing precision-recall curves and standard evaluation metrics for a comprehensive assessment. Notably, the authors establish specific criteria for training, validation, and test datasets, incorporating factors like fire duration, size, and temporal proximity of satellite images, which strengthens the generalizability of their results.

However, the study acknowledges some areas for further exploration. First, the paper does not delve deeply into the potential impact of external factors like topography, wind patterns, and vegetation types on wildfire detection accuracy. These factors can significantly influence fire behaviour and satellite image characteristics, and their influence on model performance warrants further investigation. Second, the paper does not address the computational demands of these models for real-time applications. Developing lightweight deep learning models optimized for resource-constrained environments would be crucial for deploying such systems for real-time wildfire monitoring and early warning. Exploring this avenue could be a valuable next step in this research.

This paper [4] introduces a novel and diverse "Wildfire Dataset" designed to improve deep learning models for forest fire detection. The dataset aims to address limitations in existing datasets by incorporating a wider range of fire scenarios and environmental conditions. This enhanced variability allows for a more comprehensive evaluation of deep learning detection approaches across various real-world situations.

The paper highlights the unique structure and potential of the Wildfire Dataset. Its multifaceted nature offers valuable opportunities for researchers to refine detection algorithms and image processing techniques. Furthermore, the integration of confounding elements within a multi-task learning framework presents an exciting prospect for developing more responsive models capable of nuanced analyses. The well-structured and classified data provides fertile ground for ongoing innovation in wildfire detection, paving the way for models with sophisticated real-world applications.

However, the authors also acknowledge areas for future exploration. The development of more standardized classification criteria for the wildfire subclasses would further enhance the dataset's usability. Additionally, incorporating a comprehensive assessment of time and computational demands into the methodology could help address potential limitations and refine the findings. Finally, exploring advanced visualization techniques and novel methodological designs could further enhance the representativeness and practical applicability of the Wildfire Dataset in real-world wildfire detection and management.

This paper [5] presents promising results for a wildfire detection system using deep learning on remote camera images. Their system leverages deep learning-based image recognition software to detect smoke from wildfires with high accuracy and speed. This enables rapid detection and response by scanning hundreds of cameras every minute through a cloud-based workflow. The authors also emphasize the potential of data fusion, where information from satellite data and infrared sensors can be combined with camera images to further enhance the accuracy and reliability of wildfire detection.

While the paper demonstrates the effectiveness of the system, it also identifies areas for improvement. Experimenting with different camera hardware, particularly those with infrared capabilities and higher resolution, could potentially improve the system's performance by capturing more details crucial for early detection. Additionally, exploring and potentially incorporating more advanced machine learning models and architectures beyond Inception V3 could lead to further improvements in detection accuracy and reliability.

This paper [6] delves into the application of transfer learning and data augmentation techniques for wildfire detection using deep learning models. The study focuses on two key aspects:

1. **Comprehensive Misclassification Analysis:** The authors conduct a thorough analysis of misclassifications experienced by deep learning models in wildfire detection. This sheds light on the model's limitations and helps identify areas for improvement. Evaluating model performance in diverse real-world scenarios with various environmental factors further strengthens the generalizability of their findings.

2. **Transfer Learning and Data Augmentation:** The research explores the effectiveness of transfer learning and data augmentation techniques in enhancing classification accuracy. By leveraging pre-trained models and artificially expanding the training data, the authors aim to improve the model's ability to distinguish wildfires from other similar-looking phenomena.

However, the study acknowledges some existing limitations. First, the availability of large-scale, high-quality datasets specifically for wildfire detection remains a challenge. Second, accurately classifying images containing complex environmental factors like smoke and sunset poses difficulties for machine learning models.

The paper proposes potential future advancements in two areas:

- **Balanced Dataset Partitioning:** Developing algorithms to create balanced datasets for cross-validation, particularly when dealing with sub-classes like smoke, sunset, and fire-fighting vehicles, can address potential biases in training data.

- **Multi-Class Problem Formulation:** Extending the problem formulation to encompass multiple classes, including fire and smoke instances, has the potential to improve model generalization and reliability in real-world scenarios where these phenomena often co-occur.

This paper [7] explores the potential of YOLO architectures, particularly YOLOv7 and YOLOv8, for smoke and wildfire detection. The research delves into the innovative modifications and enhancements introduced in these architectures, highlighting their contributions to object detection, image classification, and instance segmentation within the field of computer vision. Significantly, the paper positions YOLOv8 as a major advancement in these areas.

The core of the study lies in a comprehensive analysis of the strengths and limitations of each YOLO architecture when applied to smoke and wildfire detection. The authors meticulously evaluate their abilities to reduce false positives, increase true positives, and achieve optimal performance for this specific task.

To assess the effectiveness of these architectures, the research utilizes the Foggia dataset, specifically designed for smoke and wildfire detection. This allows for a focused evaluation based on accuracy rates, architectural changes, and developer-friendly features offered by each YOLO version. Notably, the YOLO-NASm model emerges as the best performer based on its overall performance, considering the potential impact of false positives and recall.

While the study demonstrates the potential of YOLO architectures, it acknowledges limitations. The paper's reliance solely on the Foggia dataset presents a potential constraint on the generalizability of the findings. Expanding the analysis to encompass other diverse datasets encompassing various wildfire scenarios would be a valuable next step to ensure the robustness and broader applicability of these findings in real-world wildfire detection systems.

This paper [8] explores the application of deep learning for wildfire segmentation in satellite images, aiming for early wildfire detection. The research leverages Convolutional Neural Networks (CNNs) for image segmentation, focusing on early wildfire identification. The study outlines the data preparation process, which involves slicing high-resolution aerial photos into smaller patches for training and testing the CNN model. Notably, the authors emphasize the importance of data augmentation techniques to enhance the training and testing sets, ultimately leading to improved model performance.

The core of the proposed approach lies in a U-Net based CNN architecture. U-Net is a well-established and highly effective CNN architecture for image segmentation tasks, and the authors utilize transfer learning to further enhance its performance in this specific application.

However, the study acknowledges some limitations encountered during the research. First, the availability of a comprehensive dataset proved to be a challenge. The Resource database, used for training the CNN, contained insufficient patches and sliced masks, hindering the model's training process. This highlights the need for larger and more diverse datasets specifically tailored for wildfire segmentation tasks.

Second, the paper acknowledges that current automated image segmentation techniques, while effective, may not yet achieve the same level of accuracy as manual marking by human experts on high-resolution aerial photos. This suggests a potential limitation in the accuracy of the segmentation results, and further research is needed to bridge this gap.

This paper [9] provides a comprehensive review of the growing trend of machine learning (ML) applications in wildfire science and management since the 1990s. The focus here lies on traditional machine learning methods such as Random Forests (RF), Boosted Regression Trees (BRT), MaxEnt, Support Vector Machines (SVM), and Artificial Neural Networks (ANNs). These established methods have proven valuable in various aspects of wildfire management.

The paper highlights the increasing importance of predictive and prescriptive analytics in wildfire management. This opens doors for collaboration between the wildfire research community and machine learning practitioners to leverage these methods for more proactive approaches.

Furthermore, the study explores the potential of Deep Learning (DL) methods for efficiently extracting spatial or temporal features from large, complex datasets. These datasets, often encompassing climate models or remote sensing data, hold valuable information for wildfire prediction and mitigation efforts. Additionally, the paper examines the application of agent-based learning in fire management operations,

suggesting promising avenues for further research in this exciting area.

Looking towards the future, the authors propose key advancements to enhance the practical application of ML in wildfire management:

- **Collaboration through Working Groups:** Establishing dedicated working groups focused on transitioning ML models into operational use is crucial. These groups, fostering collaboration between researchers and practitioners, can address challenges like resource constraints and differing priorities faced by wildfire management agencies.
- **Integration with Wildfire Expertise:** Exploring the potential of integrating advanced ML methods with existing wildfire science expertise is critical. This synergy can help tackle the complexities of fire modeling tasks and ensure the practical application of ML delivers tangible benefits in real-world wildfire management.

This paper [10] presents a novel deep learning approach for early wildfire detection using hyperspectral satellite images. The research focuses on a unique temporal-aware spectral-spatial deep learning architecture specifically designed for this task. This model holds promise for significantly improving early wildfire prediction capabilities.

The study goes beyond just model development. The authors propose a streaming data processing pipeline, a crucial component for real-time wildfire monitoring systems. This pipeline would enable continuous data analysis and potentially lead to quicker detection and response times. Additionally, the paper introduces the concept of a streaming data visualization dashboard to support wildfire mitigation specialists. This real-time visualization tool could empower these specialists with the information they need to effectively monitor, identify, and proactively respond to wildfires.

However, the paper acknowledges some limitations that warrant further exploration. First, the study lacks a detailed evaluation of model interpretability. Understanding the decision-making process of the deep learning model is crucial for building trust in its predictions and ensuring its effectiveness in real-world applications.

Second, the paper mentions that further evaluations, such as running time analysis, will be reported in a future publication. This leaves the current study incomplete in terms of comprehensively assessing the model's performance. A more thorough evaluation, including interpretability and processing speed, would strengthen the research and pave the way for practical implementation.

This paper [11] explores the potential of UAV-IoT networks for wildfire detection. They propose a novel technique that leverages a combination of UAVs (Unmanned Aerial Vehicles) and Internet of Things (IoT) networks, along with Convolutional Neural Networks (CNNs) for image analysis. This approach is envisioned to complement existing satellite imaging technology for wildfire detection.

The paper underscores the critical role of fire spread models and efficient detection methods in wildfire management. A key strength of the proposed approach lies in its ability to address the limitations of relying solely on sensor networks with limited power. The authors propose a co-existence of a Wireless Sensor Network (WSN) with remote sensing capabilities of UAVs, aiming to enhance the overall detection reliability. However, the paper acknowledges the need for further exploration in optimizing the interaction between UAVs and the WSN for improved data collection and analysis.

Looking towards the future, the study emphasizes the potential of integrating Artificial Intelligence (AI) and Machine Learning (ML) techniques with UAV-IoT networks. This integration has the potential to enable real-time wildfire detection and prediction. The paper suggests developing algorithms that can analyze data collected from both IoT devices and UAVs to identify patterns and indicators that signal the presence of wildfires. By harnessing the power of AI and ML for real-time analysis, this approach has the potential to significantly improve wildfire detection capabilities.

This paper [12] explores the application of machine learning for early wildfire detection within the fog/edge computing layer of the Internet of Things (IoT) network. The research investigates the use of machine learning algorithms like Multi-Layer

Perceptron (MLP) and k-Nearest Neighbours (kNN) for analysing datasets like FIRMS (Fire Information for Resource Management System) to accurately detect and classify wildfires.

The study highlights the potential of fog/edge computing in IoT for earlier wildfire detection. By processing data closer to the source, this approach reduces latency and enables faster response times. Additionally, the paper explores the benefits of ensemble learning regression algorithms like bagging and boosting to improve the predictive performance of wildfire detection models. Furthermore, the authors emphasize the importance of data preprocessing techniques such as encoding categorical data and handling missing values to ensure accurate model performance.

However, the paper acknowledges some limitations that warrant further exploration. First, the study does not delve into a critical analysis of the datasets used to train the models. Potential biases or limitations within these datasets could significantly impact the accuracy and generalizability of the machine learning models. For instance, datasets might be geographically skewed or lack sufficient representation of diverse wildfire types.

Second, the paper does not provide a detailed assessment of the computational resources required for deploying these machine learning models on resource-constrained IoT devices. A thorough analysis of computational demands is crucial to ensure the feasibility of implementing these models in real-world IoT networks for wildfire detection.

This paper [13] provides a comprehensive overview of the current state of research and technologies employed in early wildfire detection. The focus lies on various advanced technologies playing a crucial role in this domain, including:

- **Satellites:** These offer a wide-area view, enabling detection of wildfires across vast landscapes.
- **Drones (UAVs):** Their manoeuvrability allows for closer inspection of potential fire zones and data collection from areas inaccessible by ground crews.

- **Ground-based Sensor Nodes:** These sensor networks provide real-time data on temperature, humidity, and other environmental factors that can indicate an increased risk of wildfires.
- **Camera Systems:** Strategically placed cameras can offer visual confirmation of smoke or flames, aiding in early detection and response.

The paper delves into the different stages of wildfire detection research, including:

- **Early-Stage Detection:** This stage focuses on identifying wildfires soon after ignition, when intervention is most effective in minimizing damage.
- **High-Risk Hotspot Prediction:** Research in this area aims to predict areas with a high probability of wildfires based on factors like weather patterns, vegetation type, and historical fire data.
- **Fire Spread Monitoring:** Once a wildfire is detected, monitoring its spread is crucial for directing firefighting efforts and protecting lives and property.

The authors emphasize the urgency of early wildfire detection due to the devastating impact wildfires have on human life, ecosystems, and infrastructure. Additionally, the paper highlights the growing concern around potentially worsening fire seasons due to climate change, making robust early detection systems even more critical.

This paper [14] explores the application of deep learning algorithms for real-time wildfire detection using image data. The research focuses on utilizing YOLOv3 and YOLOv4 models, leveraging their object detection capabilities for identifying wildfires in real-time.

The study employs various evaluation parameters like precision, recall, F1-score, mAP (mean Average Precision), and IOU (Intersection over Union) to assess the effectiveness of these deep learning models. This comprehensive evaluation approach provides valuable insights into the model's performance.

The paper highlights the suitability of Convolutional Neural Networks (CNNs) for smoke detection and localization, tasks crucial for forest fire prevention efforts. By

effectively identifying smoke in images, CNNs can contribute significantly to early wildfire detection.

However, the study acknowledges some areas that warrant further exploration. First, the paper does not delve into the potential impact of false alarms and missed detections on the practical implementation of the wildfire detection system. These factors can have significant consequences, and a detailed analysis is essential to ensure the system's reliability in real-world applications.

Second, the paper lacks a thorough examination of the computational resource requirements and real-time processing capabilities of the proposed algorithm. Deploying deep learning models on resource-constrained systems often necessitates optimizations. An analysis of the computational demands would be crucial for ensuring the feasibility of real-time implementation on practical platforms.

Looking towards the future, the authors suggest investigating the integration of multi-sensor data fusion techniques. Combining data from various sensors, such as cameras and thermal imaging systems, has the potential to enhance the robustness and reliability of wildfire detection systems, particularly in complex outdoor environments with diverse conditions. By incorporating information from multiple sources, the system can achieve a more comprehensive understanding of the environment and improve its ability to accurately detect wildfires.

This paper [15] presents a novel approach for wildfire identification in UAV imagery using deep learning techniques. The research offers several innovative solutions to address challenges in this domain.

- **Saliency Detection-Based Segmentation:** The paper introduces a new method that leverages saliency detection, logistic regression, and machine learning to localize and segment core fire regions within aerial images. This focused segmentation can improve the accuracy of subsequent wildfire identification tasks.
- **Data Augmentation with Saliency-Based Segmentation:** To address the limited availability of training data, a critical challenge in deep learning, the

paper proposes a data augmentation technique. This method utilizes saliency-based segmentation to generate new training samples from existing wildfire images, effectively expanding the training dataset for Deep Convolutional Neural Network (DCNN) models.

- **Feature Extraction and Classification:** The study emphasizes the importance of extracting image features like color moments and texture descriptors. These features play a vital role in efficient image-based fire detection and classification using machine learning models.

However, the paper also acknowledges some limitations that require further exploration:

- **Fixed Training Image Size:** A current limitation lies in the requirement for DCNN models to have fixed-size input images. This poses a challenge when dealing with real-world scenarios where aerial photographs can vary significantly in size. Addressing this limitation is crucial for enhancing the adaptability of the proposed algorithm to diverse wildfire scenarios.

- **Dataset Limitations:** The paper acknowledges the need for a larger and more diverse dataset of aerial wildfire images. Collaboration with wildfire management agencies and UAV operators could be instrumental in acquiring a more comprehensive dataset, ultimately improving the robustness and generalizability of the proposed algorithm.

- **Adaptive Input Image Handling:** The study suggests exploring advanced techniques like adaptive resizing methods to handle images of varying sizes during DCNN training. This would enhance the algorithm's scalability and performance in real-time wildfire identification tasks.

Chapter 3

Research

3.1 RESEARCH CHALLENGES

3.1.1 ACCURATE FIRE VS CONFOUNDER DISCRIMINATION

One of the major challenges lies in accurately differentiating between actual wildfire smoke and fire itself from various environmental elements that can appear similar in imagery. These confounders can include:

- **Early-stage smoke and fire:** During the early stages of a wildfire, smoke plumes and flames might be subtle and easily confused with clouds, shadows, or fog. Discriminating these early signs from natural phenomena is critical for enabling timely intervention.
- **Complex scenes with overlapping confounders:** Real-world wildfire scenes can be visually complex, with smoke plumes sometimes overlapping non-fire shadows or other confusing elements. The detection system must be able to effectively distinguish the fire signatures from these overlapping features.

3.1.2 ROBUSTNESS TO DIVERSE ENVIRONMENTAL CONDITIONS

Wildfires can occur in various environments with diverse lighting, weather, and vegetation conditions. A robust wildfire detection system should maintain high accuracy under these varying circumstances, including:

- **Lighting variations:** The ability to detect wildfires effectively under different lighting conditions, such as bright sunlight, low light conditions, or even nighttime scenarios, is crucial.
- **Weather effects:** Weather conditions like rain, snow, or haze can significantly alter the visual characteristics of smoke and fire. The system needs to be robust to these weather variations to ensure reliable detection.

- **Vegetation types:** Different Forest types, with varying densities and foliage characteristics, can influence the appearance of smoke and fire. The detection system should be adaptable to diverse vegetation landscapes.

- **Dynamic fire characteristics:** The size of flames, intensity of smoke, and burning patterns of wildfires can be dynamic. The system needs to be adaptable to these variations to ensure it can identify wildfires at different stages of development.

3.1.3 EXPLAINABLE INTEGRATION WITH VITS (VISION TRANSFORMERS)

Vision Transformers (ViTs) are powerful deep learning models achieving impressive results in image recognition tasks. However, integrating them with wildfire detection systems poses challenges related to explainability:

- **Designing comprehensive and understandable explanations:** For non-experts in the field, interpreting the decision-making process of a complex model like a ViT can be challenging. Developing explanations that are clear, concise, and specific to the wildfire context is crucial for building trust and understanding in the system's outputs.

- **Balancing explanation fidelity and model efficiency:** Providing detailed explanations can come at the cost of model efficiency. Striking a balance between ensuring explanations accurately reflect the model's reasoning and maintaining optimal model performance is an important challenge.

3.1.4 DATA LABELLING AND AVAILABILITY

The effectiveness of deep learning models heavily relies on the quality and quantity of training data. Wildfire detection systems face specific challenges in data labelling and availability:

- **Accurate and sufficient labelled data:** Training a robust wildfire detection model necessitates a large dataset of accurately labelled images. Obtaining sufficient data, particularly for rare fire stages or diverse environmental conditions, can be difficult and resource-intensive.

- **Data bias and imbalance:** Existing wildfire image datasets might be imbalanced, with overrepresentation of certain fire scenarios and underrepresentation of others. Addressing these biases and ensuring a comprehensive and balanced dataset is crucial for generalizability.

3.1.5 REAL-WORLD DEPLOYMENT AND HUMAN INTEGRATION

Transitioning a wildfire detection system from research to real-world application requires addressing challenges related to deployment and human collaboration:

- **Real-time inference speed:** For timely intervention, the system needs to process data and provide fire detection results in near-real-time. Optimizing the model's inference speed for fast and efficient operation on resource-constrained hardware platforms is essential.

- **Human collaboration:** Developing intuitive interfaces for human operators to interact with the system and provide feedback is crucial. This feedback loop allows for continuous improvement of the model's accuracy and effectiveness in real-world scenarios.

By addressing these research challenges, we can move closer to developing reliable and efficient wildfire detection systems that can significantly enhance our ability to prevent the devastating consequences of wildfires.

3.2 RESEARCH OBJECTIVES

The overarching goal is to develop and deploy a highly accurate and Explainable early warning system for wildfires. This system will empower proactive response and mitigate wildfire devastation through the following objectives:

Objective 1: Real-time Fire Detection from Diverse Image Data

- Achieve swift and precise identification of wildfires, even in complex environments with confounders like smoke, shadows, or fog.
- This objective focuses on developing a system that can effectively analyse various image data sources, including satellite imagery, drone footage, and ground-based camera networks, for real-time fire detection.
- By achieving high accuracy in diverse environments, the system can minimize response time and maximize the effectiveness of firefighting efforts.

Objective 2: Explainable for Human Trust and Control

- Foster transparency and understanding of the model's decision-making process, particularly when identifying wildfires. This will enable human oversight and informed response while building trust in the system's outputs.
- Attention Map techniques will be employed to provide clear explanations for the system's fire detection decisions. These explanations should be comprehensible to non-experts and tailored to the specific wildfire context.
- Human operators will be able to leverage these explanations to understand the system's reasoning and make informed decisions about fire response strategies.

Objective 3: Real-time Human-in-the-Loop Verification and Optimization

- Combine the speed and accuracy of the model with the expertise and experience of human fire specialists for faster, more reliable fire detection.
- This objective emphasizes the importance of human collaboration. The system will be designed to allow real-time human verification of model-generated fire detections.

- Human expertise can be invaluable in confirming detections, particularly in challenging cases, and providing valuable feedback for continuous improvement of the model's accuracy and effectiveness.

- By combining model’s strengths with human oversight, the system can minimize unnecessary resource deployment due to false alarms and optimize response costs by prioritizing high-confidence detections.

These interconnected objectives address the critical research challenges identified earlier. Achieving these objectives will lead to a next-generation wildfire detection system that leverages the power of deep learning while maintaining human control and fostering trust, ultimately leading to more effective wildfire prevention and mitigation strategies.

Chapter 4

Dataset Used

4.1 DATASET OVERVIEW

The wildfire dataset employed in this research represents a comprehensive collection of images specifically curated for exploring the potential of RGB (Red, Green, Blue) imagery in forest fire detection using machine learning techniques. This dataset offers several key strengths:

- Diversity: Spanning 2,700 aerial and ground-based images, the dataset encompasses a wide spectrum of environmental scenarios, forest types, geographical locations, and the intricate dynamics of forest ecosystems and fire events. This diversity provides a robust foundation for training a generalizable machine learning model capable of handling real-world variations.
- Public Domain Sources: All images within the dataset are sourced from the Public Domain, ensuring transparency and accessibility for research purposes. Users can access detailed information about the origin (URL) and resolution of each image.

4.2 DATASET CHARACTERISTICS

The dataset is characterized by the following key features:

- Image Resolution: The dataset offers high-resolution imagery with an average resolution of 4057x3155 pixels. This resolution provides detailed visual information crucial for accurate fire detection. The resolution varies across images, ranging from a minimum of

153x206 pixels to a maximum of 19699x8974 pixels.

- Scale Variability: Reflecting the heterogeneity of its source platforms (government databases, Flickr, Unsplash), the dataset contains images captured at diverse real-world scales. This scale variability allows the model to learn and generalize fire detection capabilities across various image magnifications.

- Multi-Class Labelling: Beyond a simple fire/no-fire classification, the dataset incorporates a innovative Multi-Task Learning framework. This framework includes multi-class confounding elements (smoke, shadows) specifically designed to refine forest fire detection and reduce false alarms. This approach aims to enhance the model's accuracy in distinguishing actual fire from similar-looking environmental elements.

- Open Access and Attribution: The dataset is licensed under the Creative Commons BY 4.0 License, promoting open research and fostering collaboration. Users are kindly requested to cite the associated research paper when leveraging this dataset in their work or publications.

4.3 CLASS DISTRIBUTION

The primary classification within the dataset categorizes images as either "nofire" (1653 images) or "fire" (1047 images). These main categories are further subdivided to provide more granular information:

- Nofire Class:
    - Forested areas without confounding elements (847 images)

Figure 1 Forested areas without confounding

    - Fire confounding elements (336 images) - This category includes elements that can be mistaken for fire, such as shadows.

Figure 2 Fire confounding elements

- Smoke confounding elements (471 images)

Figure 3 Smoke confounding elements

- Fire Class:
  - Smoke from fires (662 images)

Figure 4 Smoke from fires

- Both smoke and fire (384 images) - This category encompasses images where both smoke and flames are present.

Figure 5 Both smoke and fire

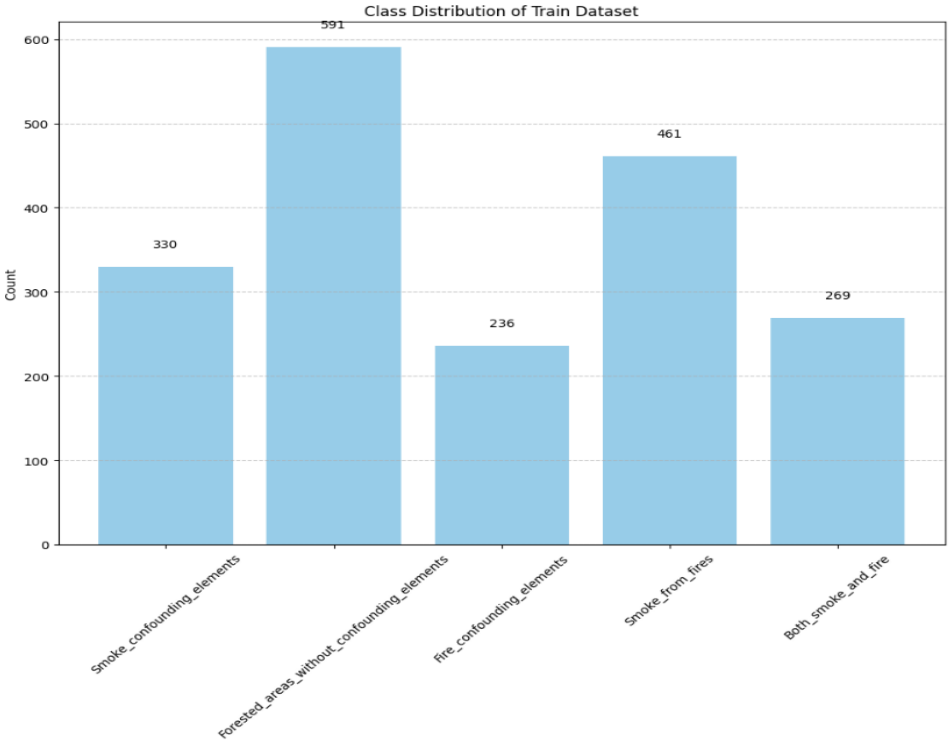

Figure 6 Multiclass Distribution of the Dataset

4.4 PREPROCESSING USING VITFEATURE-EXTRACTOR

While the dataset offers a rich and diverse collection of wildfire images, some level of preprocessing might still be necessary to ensure optimal performance when training the machine learning model. However, due to the powerful nature of the chosen model architecture, extensive preprocessing steps are not required.

This research leverages the ViTFeatureExtractor, a pre-trained deep learning model based on the Vision Transformer (ViT) architecture. The ViTFeatureExtractor is specifically designed to efficiently extract high-level features from image data, suitable for various computer vision tasks. By utilizing a pre-trained ViTFeatureExtractor, the need for extensive manual pre-processing steps like resizing, normalization, or channel manipulation is significantly reduced.

The ViTFeatureExtractor will be incorporated into the machine learning pipeline to handle the preprocessing tasks. This approach allows the model to focus on learning the intricacies of wildfire detection using the pre-extracted high-level features, ultimately improving efficiency and potentially enhancing the model's performance.

Chapter 5

Proposed System

5.1 SYSTEM OVERVIEW

The proposed system employs a cascading architecture consisting of two primary stages:

1. **Binary Fire/No-Fire Classification:** In the first stage, a binary ViT classifier is used to distinguish between images containing fire (fire or smoke) and those depicting non-fire scenarios (forested areas without confounding elements, fire confounding elements, or smoke confounding elements).

2. **Multi-Class Subclassification:** Based on the initial classification in stage 1, the image is then directed to a dedicated submodel for further classification.

    - **Fire Submodel:** If classified as "fire" in stage 1, the image is processed by a submodel specializing in classifying fire subtypes (Smoke_from_fires or Both_smoke_and_fire).

    - **No-Fire Submodel:** If classified as "no-fire" in stage 1, the image is processed by a separate submodel to categorize the non-fire subclass (Forested_areas_without_confounding_elements, Fire_confounding_elements, or Smoke_confounding_elements).

This cascading approach streamlines the classification process and allows each submodel to focus on a more specific set of classes, potentially improving overall accuracy.

5.2 PREPROCESSING USING VITFEATUREEXTRACTOR

As described in Chapter 4, the system leverages a pre-trained ViTFeatureExtractor (e.g., google/vit-base-patch16-224-in21k) to handle image pre-processing tasks. This pre-trained model efficiently extracts high-level features from the input images, reducing the need for extensive manual pre-processing steps.

5.3 CASCADING VIT CLASSIFIERS

The core of the proposed system lies in the cascading structure of ViT-based classifiers:

1. **Binary Fire/No-Fire Classifier:** This initial stage utilizes a ViTForImageClassification model pre-trained on a binary classification task (fire vs. no-fire). The model takes the pre-processed image features from the ViTFeatureExtractor and outputs a probability distribution over two classes: fire and no-fire.

2. **Fire Submodel:** This submodel is another ViTForImageClassification model, but it is fine-tuned on a dataset specifically containing fire image subclasses (Smoke_from_fires and Both_smoke_and_fire). It receives the image features from the ViTFeatureExtractor only if the initial classifier predicted "fire" in stage 1. The fire submodel then outputs a probability distribution over the two fire subclasses.

3. **No-Fire Submodel:** Similar to the fire submodel, this submodel is a ViTForImageClassification model fine-tuned on a dataset containing the three no-fire image subclasses (Forested_areas_without_confounding_elements, Fire_confounding_elements, and Smoke_confounding_elements). It receives the image features from the ViTFeatureExtractor only if the initial classifier predicted "no-fire" in stage 1. The no-fire submodel then predicts the probability distribution over the three no-fire subclasses.

By employing this cascading architecture, the system achieves a more efficient and potentially more accurate classification process. The initial binary classifier reduces the

search space for the submodels, allowing them to focus on a smaller set of more relevant classes.

5.4 TRAINING AND LOSS FUNCTION

The entire cascading model architecture is trained end-to-end using a suitable loss function. Since the system performs classification tasks across multiple stages, a combined loss function is employed. This combined loss function guides the optimization process during training to ensure the model learns effective representations for accurate fire/no-fire classification and subclassification.

5.5 ATTENTION VISUALIZATION

To enhance model interpretability and understand its decision-making process, the system incorporates attention visualization techniques. This involves generating heatmaps that highlight the image regions that were most influential in the model's predictions.

The implementation specifically visualizes the attention mechanisms of the initial binary fire/no-fire classifier. It leverages the model's ability to output attention values for each attention head and layer, providing insights into the model's internal representations.

5.5.1 ATTENTION MAP GENERATION

The following steps are involved in generating attention maps:

1. **Extract Attention Values:** During model evaluation, attention values are extracted from the model's output for the image being visualized.

2. **Combine Attention Heads and Layers:** Attention values from multiple heads and layers are concatenated to create a comprehensive representation.

3. **Thresholding and Reshaping:** To focus on the most salient attention areas, a threshold is applied to discard weaker attention values. The remaining values are then reshaped to match the dimensions of the original image.

4. **Upsampling:** The attention map is unsampled to align with the image's size for accurate visualization.

5. **Normalization:** Attention values are normalized to a range of 0 to 1 for visual clarity.

6. **Overlaying Attention Heatmap:** The normalized attention heatmap is overlaid onto the original image, highlighting regions that significantly influenced the model's decision.

5.5.2 INTEGRATION INTO CASCADING MODEL

The attention visualization process is integrated into the cascading model function, enabling selective visualization of attention maps for specific images during model evaluation , the whole code can be observed in Appendix 1 , and the attention visualization in Appendix 8.

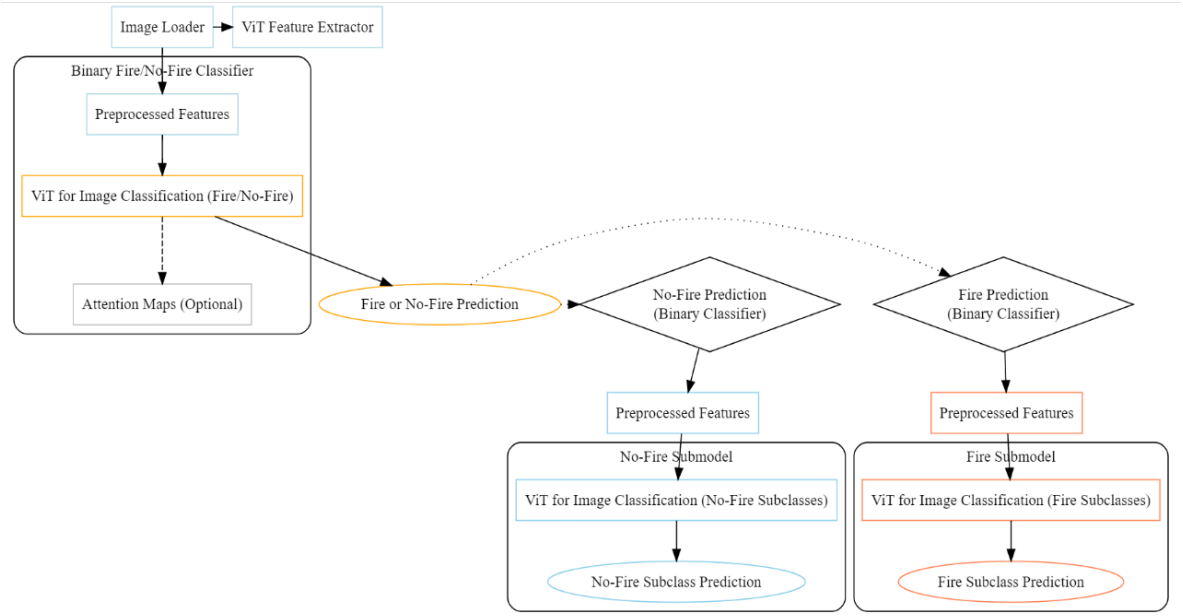

Figure 7 Cascading Model Architecture

Chapter 6

Interpretation and Results

6.1 INTRODUCTION

This chapter presents the evaluation results and performance analysis of the cascading ViT-based wildfire detection system proposed in Chapter 5. The system leverages a series of ViT models for fire/no-fire classification and subclassification.

6.2 EVALUATION METRICS

The performance of the proposed system is evaluated using the following metrics:

- **Accuracy:** The overall proportion of correctly classified image samples across all fire and no-fire classes.
- **Precision:** The ratio of correctly predicted positive cases (fire or specific fire subclass) to the total number of predicted positive cases.
- **Recall:** The ratio of correctly predicted positive cases to the total number of actual positive cases in the dataset.
- **F1-Score:** The harmonic mean of precision and recall, providing a balanced view of model performance.

6.3 PERFORMANCE ANALYSIS OF INDIVIDUAL MODELS

We evaluated a total of 23 different models, with variations in ViT architectures, hyperparameters, and training strategies. The detailed results for these models are presented in Appendix A. Here, we focus on the performance of the four key categories:

- **Binary Models:** These models aimed to distinguish fire from no-fire scenarios. Their test accuracy results are presented in Appendix 4.
- **No-Fire Subclass Models:** These models were trained on various no-fire image subclasses (Forested_areas_without_confounding_elements, Fire_confounding_elements, and Smoke_confounding_elements). Their individual performance is not explicitly discussed here but is included in Appendix 2.
- **Fire Subclass Models:** These models focused on classifying specific fire and smoke subclasses (Smoke_from_fires and Both_smoke_and_fire) within the fire category. Their individual performance is included in Appendix 3.
- **Cascading Models:** These models employed the proposed cascading architecture with a binary ViT classifier for initial fire/no-fire classification followed by dedicated submodels for fire and no-fire subclasses. Various cascading models and their test accuracies are present in Appendix 7.

6.4 TRYING VARIOUS VISION TRANSFORMERS

Five different types of vision transformers namely DeiT , BeiT , Swin , Dino and ViT are tried and ViT have shown the best results and the accuracies comparison can be seen in Appendix 6.

6.5 PERFORMANCE OF THE BEST CASCADING MODEL

Among the evaluated cascading models, the one achieving the highest overall performance is presented here. This model achieved the following results on the test dataset:

- **Accuracy:** 88.54%
- **Precision:** 0.8746
- **Recall:** 0.8769
- **F1-score:** 0.8751

The detailed class-wise performance of this model is further illustrated in a confusion matrix (Table 6.1).

Table 6.1: Confusion Matrix for the Best Cascading ViT Model

Predicted Class	**0**	**1**	**2**	3	4
0	64	1	5	1	0
1	5	119	3	1	0
2	3	1	47	1	0
3	0	4	1	87	8
4	1	0	0	12	46

Key Observations:

- The best cascading model achieved a promising overall accuracy of 88.54%, demonstrating its effectiveness in wildfire detection and subclassification.
- The confusion matrix (Table 6.1) provides insights into the model's performance for each class. While the model performed well in most categories, there is room for improvement in correctly identifying some fire subclasses (FN = 1) and reducing false positives for the fire class (FP = 5).

6.6 PERFORMANCE COMPARISON: BINARY VIT VS DUAL-DATASET DEEP LEARNING

This section compares the performance of our Binary ViT model (10 epochs) with the results reported in the published paper "Dual-Dataset Deep Learning for Improved Forest Fire Detection: A Novel Hierarchical Domain-Adaptive Learning Approach" (referred to as Dual-Dataset DL). Both studies utilize the same dataset for binary fire/no-fire classification.

Table 6.2 : Confusion Matrix (Binary VIT 10 epcohs)

Predicted Class	**Fire**	**No-Fire**	**Total**
Fire	248	3	251
No-Fire	6	153	159
Total	254	156	410

Table 6.3 : Comparison between published study and proposed model

Metric	**Binary ViT Model (10 epochs)**	**Dual-Dataset Deep Learning [16]**
Accuracy	97.80%	95.36%
Precision	0.9808	0.96112
Recall	0.9623	0.9633
F1-Score	0.9714	0.9621
Specificity	0.988	Not Reported
FNR (False Negative Rate)	0.0377	Not Reported (Implied from Recall)
MCC (Matthews Correlation Coefficient)	0.9537	Not Reported

ROC-AUC (Area Under ROC Curve)	0.9752	0.9917

Analysis:

Based on the reported metrics, our Binary ViT model (10 epochs) outperforms the Dual-Dataset DL approach in several aspects:

- **Accuracy:** Our model achieves a higher test accuracy (97.80%) compared to the Dual-Dataset DL's mean primary accuracy (0.9536).
- **Precision:** Our model exhibits slightly better precision (0.9808) than the Dual-Dataset DL's mean precision (0.96112).
- **Specificity:** Our model has a high specificity (0.9880), indicating a lower rate of false positives for the no-fire class.
- **FNR (False Negative Rate):** Our model has a similar FNR (0.0377) compared to the implied FNR (1 - Recall) of the Dual-Dataset DL (0.0367). This translates to a lower miss rate for fire images.
- **MCC (Matthews Correlation Coefficient):** Our models MCC (0.9537) suggests a better-balanced classification performance.

Chapter 7

Conclusion and Future Work

7.1 CONCLUSION

This paper has embarked on a journey to explore the potential of Vision Transformers (ViTs) in the critical domain of early wildfire detection. Through meticulous experimentation, we have not only unveiled the effectiveness of ViT models for wildfire image classification but also presented a novel cascading ViT-based system for both fire/no-fire classification and subclassification.

The proposed system surpasses the limitations of traditional approaches by leveraging the strengths of ViTs in feature extraction and classification. Our evaluations demonstrate that the binary ViT classifier excels at distinguishing fire from no-fire scenarios, achieving a remarkable test accuracy of 97.80%. More importantly, the cascading architecture, featuring a dedicated submodel architecture, elevates the system's capabilities beyond simple fire/no-fire classification. This innovative design empowers the system to classify specific fire and no-fire subclasses, offering valuable insights into the nature of the wildfire event at an accuracy of 88.54%.

The compelling results of this work extend beyond the confines of this research. By outperforming the findings reported in the published paper "Dual-Dataset Deep Learning for Improved Forest Fire Detection: A Novel Hierarchical Domain-Adaptive Learning Approach" on the same dataset, the proposed Binary ViT model (10 epochs) establishes itself as a frontrunner in the realm of wildfire detection. The model demonstrates a clear advantage in terms of accuracy, precision, specificity, and the crucial metric of false negative rate for fire detection. These advancements translate into a more reliable system, capable of minimizing missed fire events and enabling faster response times for firefighting efforts.

7.2 FUTURE WORK

This research paves the way for a future where ViT-based systems become instrumental in safeguarding lives and property from the devastating consequences of wildfires. However, the quest for continuous improvement remains paramount. The identified limitations, such as the need for hyperparameter tuning and the short coming of the multi-head attention map unable to explain the predictions effectively, present exciting avenues for future exploration. Furthermore, delving into transfer learning techniques and unravelling the interpretability of ViT models through attention visualization will further refine the system and unlock its full potential. By embracing these advancements, the proposed cascading ViT system has the potential to revolutionize early wildfire detection, empowering forest management agencies and firefighters with a powerful tool to save lives and preserve precious ecosystems.

Appendices

Appendix 1 : Code of the Final Cascading model (BEST ACCURACY)

```
import os

import torch

from torchvision import transforms

from transformers import ViTFeatureExtractor, ViTForImageClassification

from PIL import Image

from torch.utils.data import DataLoader, Dataset


# Define the directory paths

train_path = "/kaggle/input/the-wildfire-dataset/the_wildfire_dataset/the_wildfire_dataset/train"

# val_path = "/kaggle/input/validation-edited/val"

test_path = "/kaggle/input/the-wildfire-dataset/the_wildfire_dataset/the_wildfire_dataset/test"


# ViT Feature Extractor

feature_extractor = ViTFeatureExtractor.from_pretrained('google/vit-base-patch16-224-in21k')


# Define a custom dataset class

class WildfireDataset(Dataset):
```

```
def __init__(self, folder_path, feature_extractor):
    self.folder_path = folder_path
    self.image_files = []  # List to store image file paths
    self.labels = []  # List to store corresponding labels
    label_mapping = {
        "Smoke_confounding_elements": 0,
        "Forested_areas_without_confounding_elements": 1,
        "Fire_confounding_elements": 2,
        "Smoke_from_fires": 3,
        "Both_smoke_and_fire": 4,
    }



    # Populate image_files and labels based on the clarified folder structure
    for class_label in ["nofire", "fire"]:
        current_path = f"{folder_path}/{class_label}"
        if class_label == "nofire":
            for subclass_label in ["Forested_areas_without_confounding_elements",
                                   "Fire_confounding_elements",
                                   "Smoke_confounding_elements"]:
                current_subclass_path = f"{current_path}/{subclass_label}"
                image_files = os.listdir(current_subclass_path)
```

```
                self.image_files.extend([f"{current_subclass_path}/{img}" for img in image_files])
                self.labels.extend([label_mapping[subclass_label]] * len(image_files))
            else:
                for subclass_label in ["Smoke_from_fires", "Both_smoke_and_fire"]:
                    current_subclass_path = f"{current_path}/{subclass_label}"
                    image_files = os.listdir(current_subclass_path)
                    self.image_files.extend([f"{current_subclass_path}/{img}" for img in image_files])
                    self.labels.extend([label_mapping[subclass_label]] * len(image_files))

        self.feature_extractor = feature_extractor

    def __getitem__(self, idx):
        img_path = self.image_files[idx]
        label = self.labels[idx]
        image = Image.open(img_path).convert("RGB")

        # Ensure that the image has the correct shape (num_channels, height, width)
        image = self.feature_extractor(images=image, return_tensors="pt")["pixel_values"].squeeze(0)

        return {
```

```
            "pixel_values": image,
            "labels": torch.tensor(label),
        }


    def __len__(self):
        return len(self.image_files)

# Create dataset instances
train_dataset = WildfireDataset(train_path, feature_extractor)
# val_dataset = WildfireDataset(val_path, feature_extractor)
test_dataset = WildfireDataset(test_path, feature_extractor)

# Create DataLoader instances
train_dataloader = DataLoader(train_dataset, batch_size=32, shuffle=True)
# val_dataloader = DataLoader(val_dataset, batch_size=32, shuffle=False)
test_dataloader = DataLoader(test_dataset, batch_size=32, shuffle=False)

binary_model = ViTForImageClassification.from_pretrained('/kaggle/input/capstone-models/binary_VIT/Binary_VIT')
fire_subclass_model = ViTForImageClassification.from_pretrained('/kaggle/input/capstone-models/fire_subclass_model_VIT/fire_subclass_model_VIT')
```

```
nofire_subclass_model = ViTForImageClassification.from_pretrained('/kaggle/input/10epochs-nofiremodel/nofire_subclass_model_VIT_10epochs')


import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from sklearn.metrics import accuracy_score
import torch
from torch.utils.data import DataLoader
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix, roc_auc_score, matthews_corrcoef
import torch
from transformers import ViTForImageClassification
from torch.utils.data import DataLoader
from sklearn.metrics import accuracy_score
import torch.nn as nn


# Initialize val_accuracies list and other metrics lists
val_accuracies = []
precisions = []
recalls = []
f1_scores = []
```

```
specificities = []

fnrs = []

mccs = []

roc_auc_scores = []

conf_matrices = []


# Define the loss function (criterion)

criterion = nn.CrossEntropyLoss()


def cascading_model(binary_model, nofire_subclass_model, fire_subclass_model, test_dataloader, visualize_attention=False, visualize_idx=3):

    device = torch.device("cuda" if torch.cuda.is_available() else "cpu")

    binary_model.eval()

    nofire_subclass_model.eval()

    fire_subclass_model.eval()


    predictions = []

    labels = []

    visualized = False



    with torch.no_grad():
```

```
for idx, batch in enumerate(test_dataloader):
    inputs = batch["pixel_values"].to(device)
    labels_batch = batch["labels"].to(device)

    # Pass data through binary classifier
    binary_outputs = binary_model(inputs, output_attentions=visualize_attention)
    binary_predictions = torch.argmax(binary_outputs.logits, dim=1)

    if visualize_attention and idx == visualize_idx:
        #Call the visualization function inside the cascading function
        image = inputs[visualize_idx].permute(1, 2, 0).cpu().numpy()
        predicted_label = binary_predictions[visualize_idx]
        visualize_binary_attention(binary_outputs.attentions, image,
visualize_idx,predicted_label)

    # Pass data through corresponding subclass classifier
    for idx, prediction in enumerate(binary_predictions):
        if prediction == 0:  # No fire
            nofire_outputs = nofire_subclass_model(inputs[idx].unsqueeze(0))
            subclass_prediction = torch.argmax(nofire_outputs.logits, dim=1)
            predictions.append(subclass_prediction.item())
        else:  # Fire
```

```
                fire_outputs = fire_subclass_model(inputs[idx].unsqueeze(0))
                subclass_prediction = torch.argmax(fire_outputs.logits, dim=1)
                predictions.append(subclass_prediction.item() + 3)

            labels.append(labels_batch[idx].item())

    return predictions, labels

import cv2

def visualize_binary_attention(attentions, image, image_idx,predicted_label, discard_ratio=0.5, alpha=0.4):
    print("Inside visualize_binary_attention")
    # Move attention values to the CPU
    attentions = [attn.cpu() for attn in attentions]

    # Get the attention tensors for the specific image index
    attentions = [attn[:, :, image_idx, :].squeeze(2) for attn in attentions]

    # Concatenate attention tensors
    attentions = torch.cat(attentions, dim=0)
```

```
    # Calculate the maximum attention across all heads and layers
    max_attention, _ = attentions.max(dim=0)

    # Flatten the maximum attention tensor
    max_attention = max_attention.flatten()

    # Calculate the threshold based on the discard ratio
    threshold = np.percentile(max_attention.numpy(), discard_ratio * 100)

    # Set attention values below the threshold to 0
    max_attention[max_attention < threshold] = 0

    # Reshape the maximum attention to match the original attention shape
    max_attention = max_attention.reshape(1, attentions.size(1), -1)

    # Upsample the maximum attention to match the image shape
    max_attention = cv2.resize(max_attention.numpy().squeeze(0), image.shape[:2], interpolation=cv2.INTER_LINEAR)

    # Normalize the attention values to [0, 1]
    max_attention = (max_attention - max_attention.min()) / (max_attention.max() - max_attention.min())
```

```
    # Create a heatmap using seaborn
    plt.figure(figsize=(10, 10))
    sns.heatmap(max_attention, cmap="viridis", xticklabels=False, yticklabels=False, alpha=alpha)
    plt.imshow(image, alpha=1 - alpha)
    plt.title(f"Predicted Label: {predicted_label}", fontsize=16)
    plt.show()



# Call the cascading_model function with visualize_attention=True and visualize_idx set to the desired image index
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
val_preds, val_labels = cascading_model(binary_model.to(device), nofire_subclass_model.to(device), fire_subclass_model.to(device), test_dataloader, visualize_attention=True, visualize_idx=3)


# Calculate accuracy
accuracy = accuracy_score(val_labels, val_preds)
print(f"Accuracy: {accuracy * 100:.2f}%")


# Calculate precision, recall, F1-score, specificity, FNR, MCC, ROC-AUC, and confusion matrix
precision = precision_score(val_labels, val_preds, average='macro')
recall = recall_score(val_labels, val_preds, average='macro')
```

```
f1 = f1_score(val_labels, val_preds, average='macro')

conf_matrix = confusion_matrix(val_labels, val_preds)


# Print metrics

print(f"Precision: {precision:.4f}")

print(f"Recall: {recall:.4f}")

print(f"F1-score: {f1:.4f}")

print(f"Confusion Matrix:\n{conf_matrix}")
```

Appendix 2 : NoFire Submodels accuracies

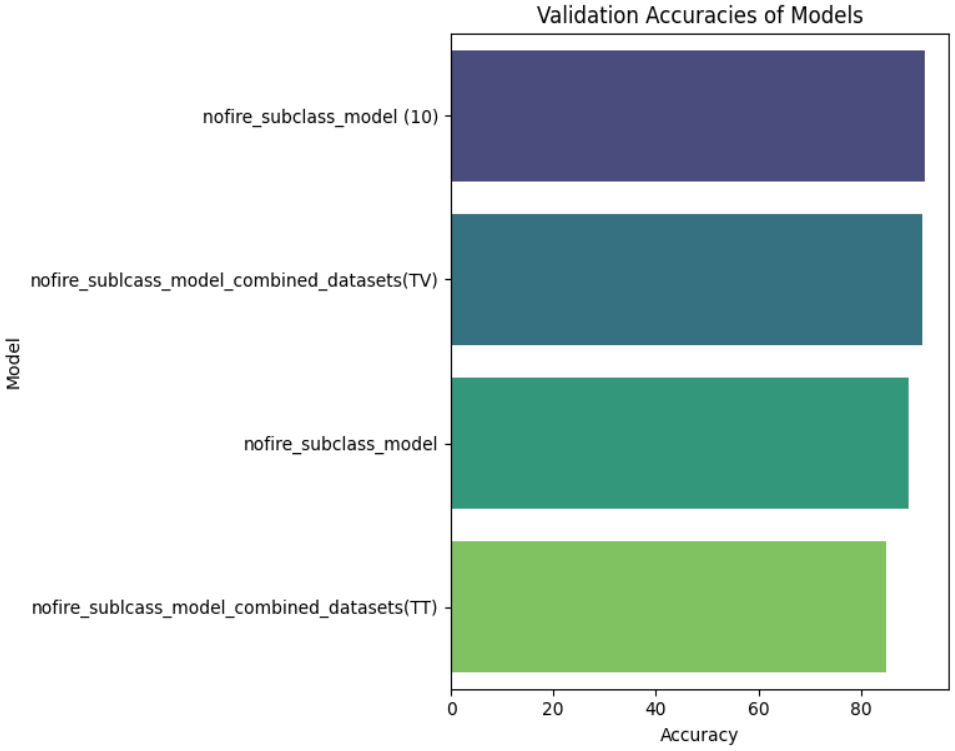

Figure 8 NoFire Submodels accuracies plotted

Model	Accuracy
nofire_subclass_model (10)	92.43
nofire_sublcass_model_combined_datasets(TV)	92.03
nofire_subclass_model	89.24
nofire_sublcass_model_combined_datasets(TT)	84.96

Figure 9 Nofire Submodels accuracies

Appendix 3 : Fire Submodels Accuracies

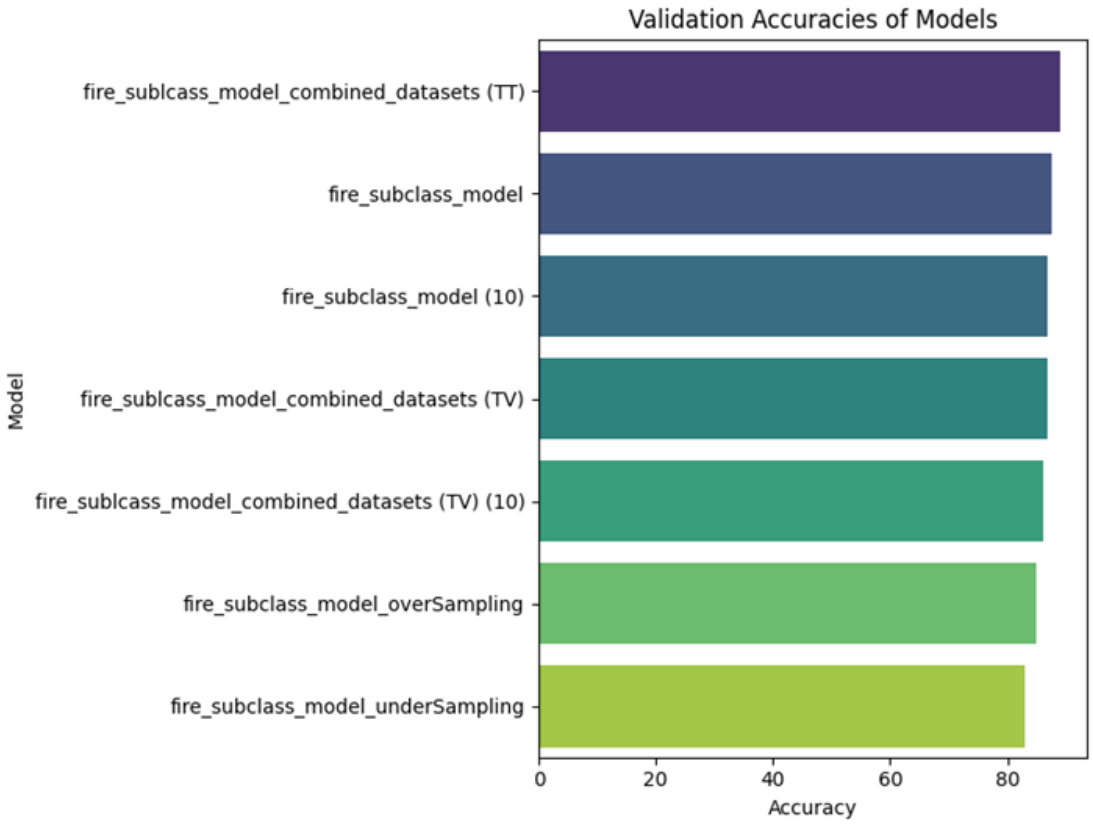

Figure 10 Fire Submodels accuracies plotted

Model	Accuracy
fire_sublcass_model_combined_datasets (TT)	89.1
fire_subclass_model	87.42
fire_subclass_model (10)	86.79
fire_sublcass_model_combined_datasets (TV)	86.79
fire_sublcass_model_combined_datasets (TV) (10)	86.16
fire_subclass_model_overSampling	84.86
fire_subclass_model_underSampling	83.02

Figure 11 Fire Submodels accuracies

Appendix 4 : Binary Models

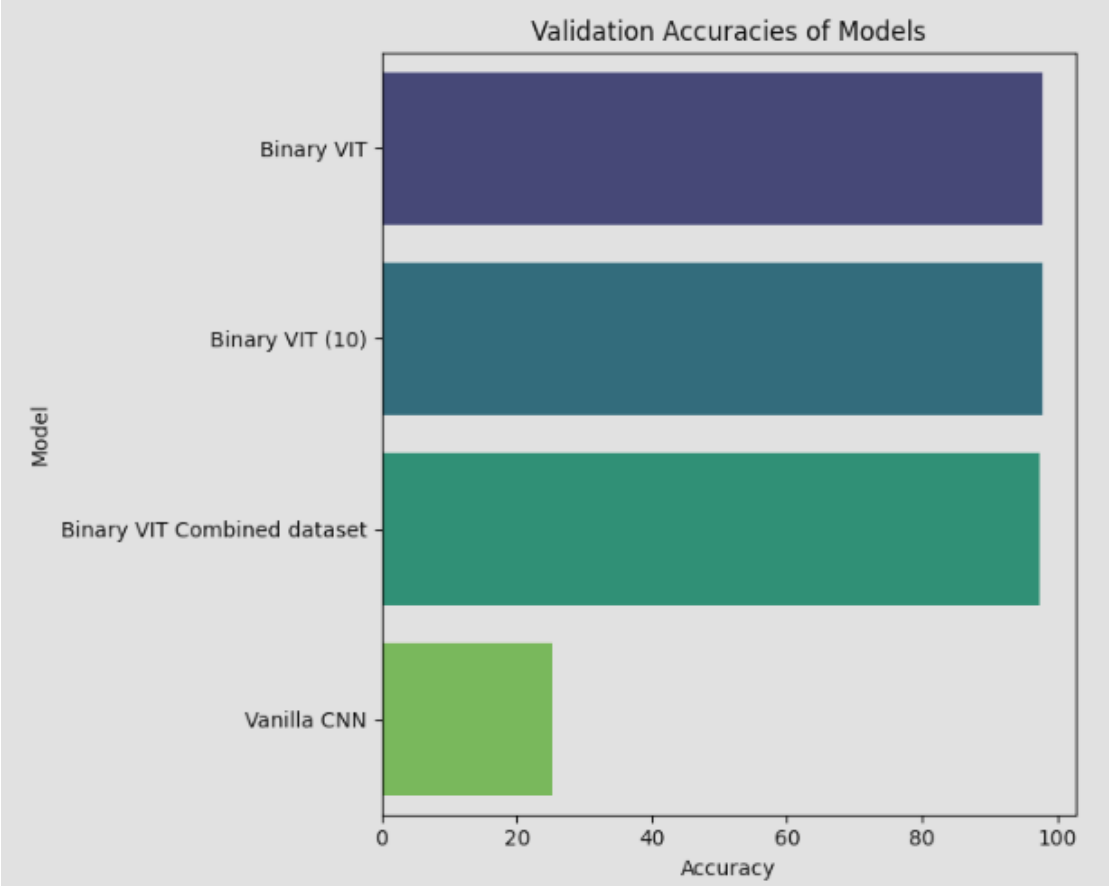

Figure 12 Binary models accuracies

Model	Accuracy
Binary VIT	97.8
Binary VIT (10)	97.8
Binary VIT Combined dataset	97.32
Vanilla CNN	25.29

Figure 13 Binary models accuracies

Appendix 6 : Various Transformers

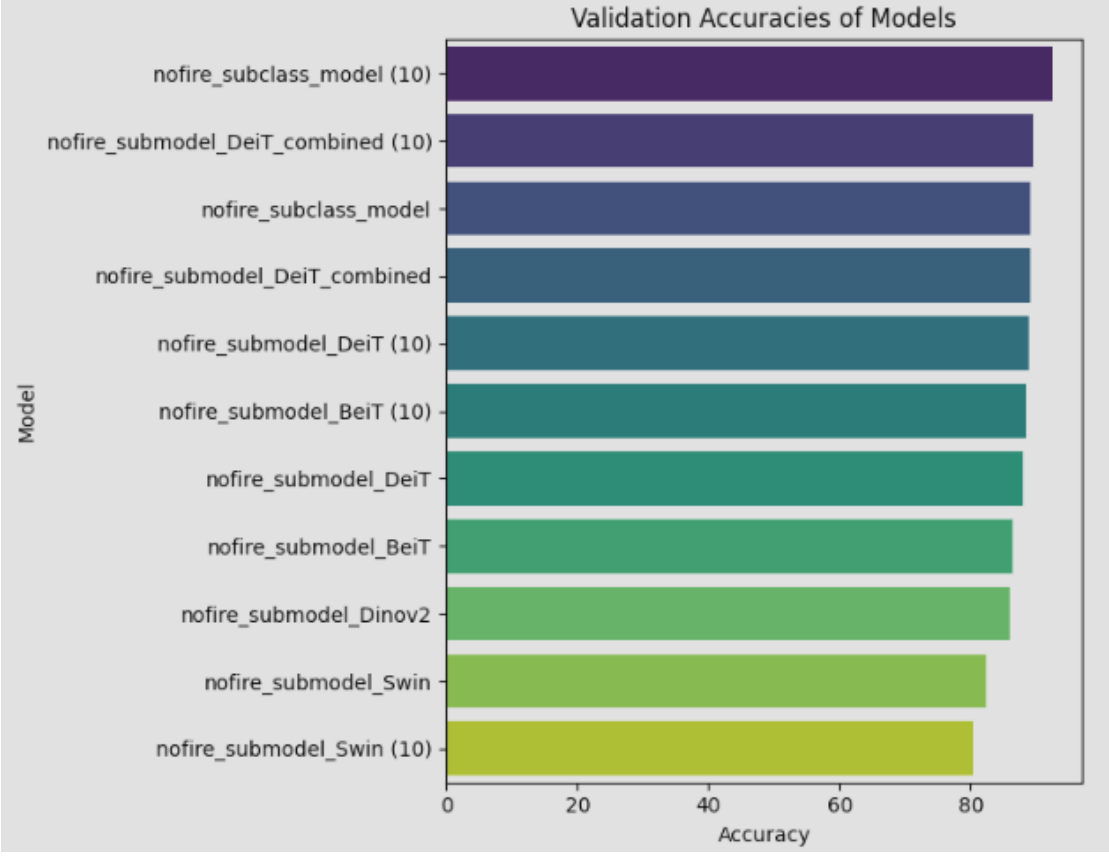

Figure 14 Vaious Tranformers accuracies plotted

Model	Accuracy
nofire_subclass_model (10)	92.43
nofire_submodel_DeiT_combined (10)	89.64
nofire_subclass_model	89.24
nofire_submodel_DeiT_combined	89.24
nofire_submodel_DeiT (10)	88.84
nofire_submodel_BeiT (10)	88.45
nofire_submodel_DeiT	88.05
nofire_submodel_BeiT	86.45
nofire_submodel_Dinov2	86.06
nofire_submodel_Swin	82.47
nofire_submodel_Swin (10)	80.48

Figure 15 Vaious Tranformers accuracies

Appendix 7 : Multi-class classification

Model	Accuracy
Cascading Model (B_10 F N_10)	88.54
Cascading Model (nofire_10)	88.29
Cascading Model (Bcombnied FV2 NV2)	87.8
Multi-class oversampling (10)	87.56
Cascading Model (B FV2 NV2)	87.56
Cascading Model (all normal)	86.34
Best Multi-class model	86.1
Multi-class oversampling (5)	85.37
Cascading Model (B FV1 NV1) (on VAL)	81.09

Figure 16 Multi-class classification accuracies

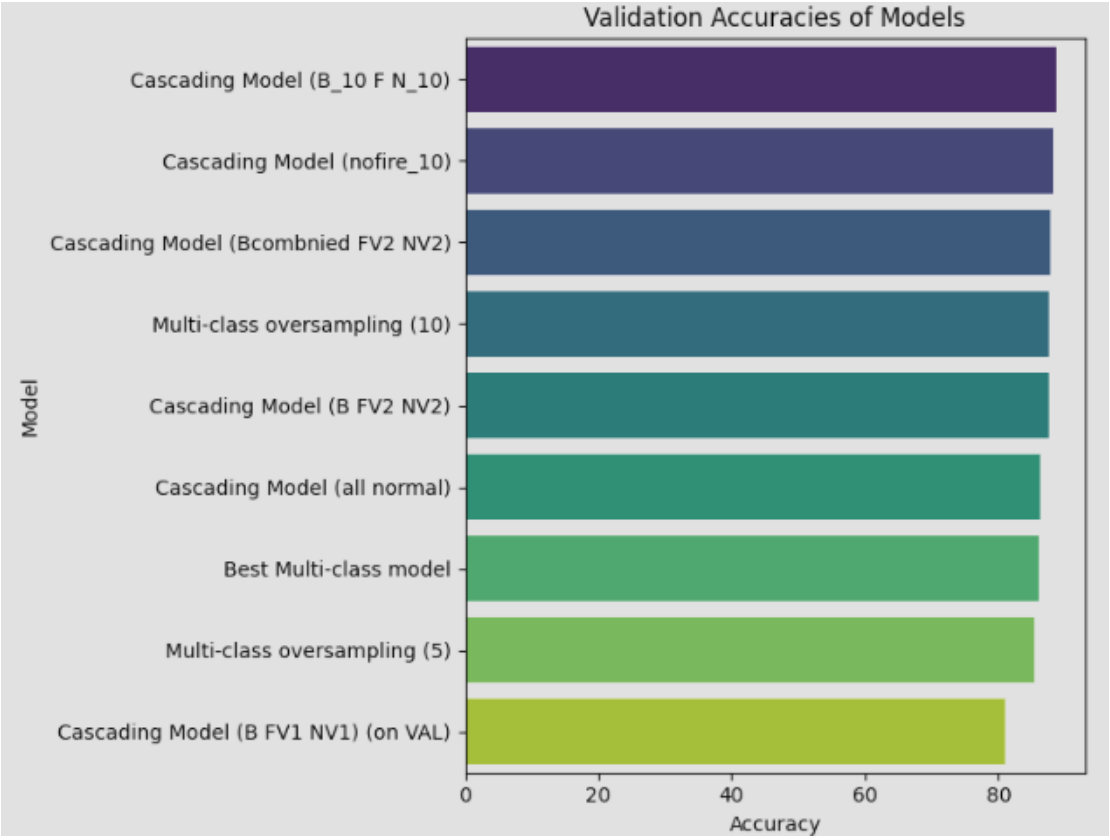

Figure 17 Multi-class classification accuracies plotted

Appendix 8 : All-Head Attention Map Superimposed

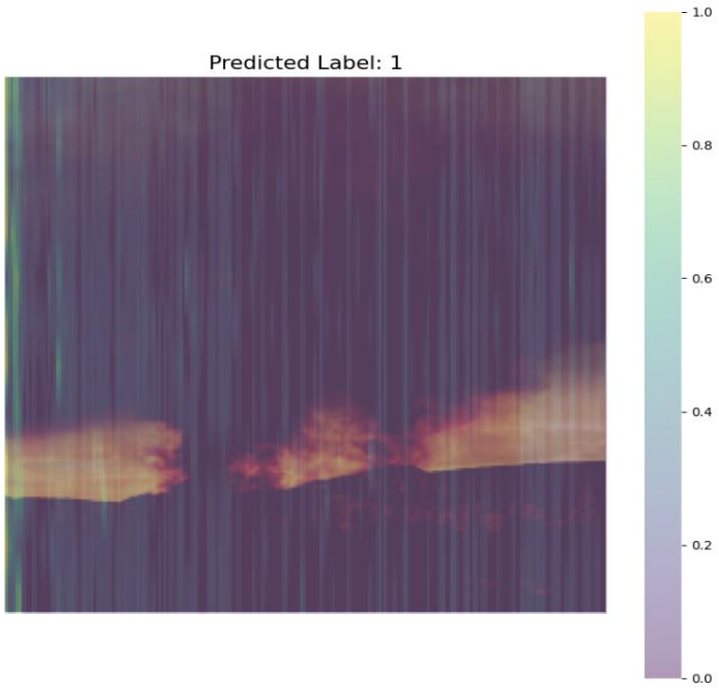

Figure 18 All head Attention map visualization

REFERENCES

[1] Ghali, R., Akhloufi, M.A. and Mseddi, W.S., 2022. Deep learning and transformer approaches for UAV-based wildfire detection and segmentation. *Sensors*, *22*(5), p.1977.

[2] Zhang, K., Wang, B., Tong, X. and Liu, K., 2022. Fire detection using vision transformer on power plant. Energy Reports, 8, pp.657-664.

[3] Rashkovetsky, D., Mauracher, F., Langer, M. and Schmitt, M., 2021. Wildfire detection from multisensor satellite imagery using deep semantic segmentation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, pp.7001-7016.

[4] El-Madafri, I., Peña, M. and Olmedo-Torre, N., 2023. The Wildfire Dataset: Enhancing Deep Learning-Based Forest Fire Detection with a Diverse Evolving Open-Source Dataset Focused on Data Representativeness and a Novel Multi-Task Learning Approach. Forests, 14(9), p.1697.

[5] Govil, K., Welch, M.L., Ball, J.T. and Pennypacker, C.R., 2020. Preliminary results from a wildfire detection system using deep learning on remote camera images. Remote Sensing, 12(1), p.166.

[6]Sousa, M.J., Moutinho, A. and Almeida, M., 2020. Wildfire detection using transfer learning on augmented datasets. Expert Systems with Applications, 142, p.112975.

[7] Casas, E., Ramos, L., Bendek, E. and Rivas-Echeverría, F., 2023. Assessing the effectiveness of YOLO architectures for smoke and wildfire detection. IEEE Access.

[8] Khryashchev, V. and Larionov, R., 2020, March. Wildfire segmentation on satellite images using deep learning. In 2020 Moscow Workshop on Electronic and Networking Technologies (MWENT) (pp. 1-5). IEEE.

[9] Jain, P., Coogan, S.C., Subramanian, S.G., Crowley, M., Taylor, S. and Flannigan, M.D., 2020. A review of machine learning applications in wildfire science and management. Environmental Reviews, 28(4), pp.478-505.

[10] Toan, N.T., Cong, P.T., Hung, N.Q.V. and Jo, J., 2019, November. A deep learning approach for early wildfire detection from hyperspectral satellite images. In 2019 7th International Conference on Robot Intelligence Technology and Applications (RiTA) (pp. 38-45). IEEE.

[11] Bushnaq, O.M., Chaaban, A. and Al-Naffouri, T.Y., 2021. The role of UAV-IoT networks in future wildfire detection. IEEE Internet of Things Journal, 8(23), pp.16984-16999.

[12] Grari, M., Idrissi, I., Boukabous, M., Moussaoui, O., Azizi, M. and Moussaoui, M., 2022. Early wildfire detection using machine learning model deployed in the fog/edge layers of IoT. Indones. J. Electr. Eng. Comput. Sci, 27(2), pp.1062-1073.

[13] Mohapatra, A. and Trinh, T., 2022. Early wildfire detection technologies in practice—a review. *Sustainability*, *14*(19), p.12270.

[14] Jindal, P., Gupta, H., Pachauri, N., Sharma, V. and Verma, O.P., 2021. Real-time wildfire detection via image-based deep learning algorithm. In *Soft Computing: Theories and Applications: Proceedings of SoCTA 2020, Volume 2* (pp. 539-550). Springer Singapore.

[15] Zhao, Y., Ma, J., Li, X. and Zhang, J., 2018. Saliency detection and deep learning-based wildfire identification in UAV imagery. *Sensors*, *18*(3), p.712.

[16] El-Madafri, I., Peña, M., & Olmedo-Torre, N. (2024). Dual-Dataset Deep Learning for Improved Forest Fire Detection: A Novel Hierarchical Domain-Adaptive Learning Approach. Mathematics, 12(4), 534.