DESIGN AND DEVELOPMENT OF AI-BASED FRAMEWORKS FOR ENVIRONMENTAL AND GEOSPATIAL DATA ANALYSIS

A Thesis Submitted
In the Partial Fulfilment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

by

Abhishek Verma
(2K22/PHDIT/01)

Under the Supervision of

Dr. Virender Ranga
&
Prof. Dinesh Kumar Vishwakarma

Delhi Technological University

Department of Information Technology

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India

February 2025

ACKNOWLEDGEMENTS

I am profoundly grateful to my esteemed PhD supervisors, **Dr. Virender Ranga** and **Prof. Dinesh Kumar Vishwakarma**, whose exceptional guidance and unwavering support have been instrumental in the completion of my PhD journey. Their exemplary discipline, unyielding focus, and relentless work ethic have inspired me and set a high standard for academic excellence. I am truly fortunate to have benefited from their profound expertise and commendable commitment throughout this challenging yet rewarding journey. Their professionalism and dedication to their craft have been a constant source of motivation, shaping my academic growth and ensuring the success of this endeavor. I thank **Mrs. Sushma Vishwakarma**, **Diya,** and **Advika** for ensuring I always had a home away from home.

No man is complete without his family, who silently toil behind the scenes to help him fight for his dreams. I thank my parents, **Mr. Surya Kumar** and **Mrs. Beena Verma,** for being the best parents one could hope for; I thank my siblings and cousins **Abhinav** and **Ranjeet, Rahul, Sidharth, Ritika, Ramnit, and Sidhantika** for doing what siblings do best, i.e., be loving in their fun way.

.Finishing PhD is a highly challenging journey, and the seniors who helped navigate this path need a special mention. To this end, I am grateful for the support of my PhD seniors, **Dr. Ankit Yadav, Dr. Deepak Dagar, Dr. Anusha Chabbra, Dr. Ananya Pandey, Dr. Ashish Bajaj, and Dr. Lakshita Agarwal.** As I went through the most challenging phase of my PhD, my juniors ensured that I never gave up and always came back stronger. I am grateful to colleagues and juniors **Bhavana Verma** and **Sajal Agarwal** for all the light-hearted conversations. I want to express my heartfelt gratitude to my dear friends **Ankit, Nity, Divyansh, Prateek, Laqshiqa,Anurag,** and **Harshit** for being an unwavering pillar of support throughout my PhD journey.

I extend my heartfelt appreciation to the state-of-the-art research lab established by my supervisor. Equipped with cutting-edge NVIDIA GPUs, it was pivotal in facilitating the success of the computationally expensive deep learning-based research experiments throughout my PhD.Finally, yet importantly, I thank God for giving me the persistence and strength to show up at my lab each day and work through the difficulties of this PhD journey.

Abhishek Verma

2K22/PHDIT/01

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I certify that the dissertation titled **"Design and Development of AI-based Frameworks for Environmental and Geospatial Data Analysis"** submitted for the Doctor of Philosophy degree is my work and has not been submitted for the award of any degree or diploma to any other University or Institute. The work done in the thesis is original and has been done by me under the supervision of my supervisors.

I also mention that the research work is original and has not been submitted by me, in part or completely, to any other University or Institution for the award of any degree or diploma.

Abhishek Verma

(Ph.D. Scholar)

Department of Information Technology,

Delhi Technological University, Delhi

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE

This is to certify that the work contained in the thesis entitled "**Design and Development of AI-based Frameworks for Environmental and Geospatial Data Analysis***"* submitted by **Abhishek Verma (2K22/PHDIT/01)** for the award of the degree of Doctor of Philosophy to Delhi Technological University, India contains original research work carried out by him under our supervision.

He has fulfilled all the requirements for the submission of the thesis as per the required standard. We hereby confirm the originality of the work and certify that the thesis has not formed the basis for the award of any degree or similar title.

Dr. Virender Ranga
(Supervisor)
Associate Professor
Department of Information Technology,
Delhi Technological University

Prof. Dinesh Kumar Vishwakarma
(Co-Supervisor)
Professor
Department of Information Technology,
Delhi Technological University

ABSTRACT

Wildfires, air pollution, and climate change are interconnected environmental challenges with far-reaching consequences for ecosystems, public health, and global climate stability. Wildfires contribute significantly to atmospheric pollution, releasing large quantities of greenhouse gases and particulate matter, accelerating ozone depletion and global warming. This warming effect causes polar ice to melt, reducing Earth's albedo and creating a feedback loop that further exacerbates climate change. Addressing these issues requires advanced monitoring and predictive systems to mitigate their impact effectively. This thesis presents the design and development of AI-based frameworks for environmental and geospatial data analysis, focusing on wildfire risk detection, air pollution prediction, and sea ice classification. The research integrates state-of-the-art deep learning models and remote sensing data to enhance the accuracy and efficiency of environmental monitoring systems. The study introduces the Swin Transformer and IGNITE-NET models for wildfire risk detection, which leverage dynamic receptive field blocks and channel fusion attention mechanisms to improve predictive accuracy while maintaining computational efficiency. These models demonstrate superior performance in classifying fire risk levels using remote sensing imagery, contributing to proactive wildfire management strategies.

In the domain of air pollution prediction, the thesis presents the BREATH-Net model, a hybrid deep learning framework that combines Bi-directional Long Short-Term Memory (BiLSTM) networks with Transformer architectures. Using satellite data, this model accurately forecasts nitrogen dioxide (NO_2) concentrations, offering a robust tool for air quality management and public health interventions. The Arctic-Net model is proposed for sea ice classification, integrating Convolutional Neural Networks (CNNs) with attention mechanisms to efficiently classify sea ice types using Synthetic Aperture Radar (SAR) images. The model outperforms existing methods in accuracy and robustness, providing valuable insights for climate research and maritime navigation. The experimental results across all three domains highlight the superior performance of the proposed models compared to traditional approaches. By combining AI with remote sensing technologies, this research contributes to developing scalable, efficient, and accurate environmental monitoring systems. The findings have significant implications for environmental policy-making, disaster management, and climate change mitigation, demonstrating the transformative potential of AI in addressing complex environmental challenges.

LIST OF PUBLICATIONS

Journal Papers

- **Abhishek Verma,** Virender Ranga, Dinesh Kumar Vishwakarma, "A novel approach for forecasting PM2.5 pollution in Delhi using CATALYST." Published in Environmental Monitoring and Assessment, Volume 196, Article number 340, (2024).
- **Abhishek Verma,** Virender Ranga, Dinesh Kumar Vishwakarma, "Arctic-Net: A Hybrid Convolutional and Attention-Based Model for Efficient Sea Ice Classification Using SAR Images ", Communicated in Super Computing.
- **Abhishek Verma,** Virender Ranga, Dinesh Kumar Vishwakarma, "IGNITE-NET: Fire Risk Prediction using Dynamic Receptive Fields and Dynamic Channel Fusion Attention." Published in Advances in Space Research (2025)(Pub: Elsevier).
- **Abhishek Verma,** Virender Ranga, Dinesh Kumar Vishwakarma, "BREATH-Net: a novel deep learning framework for NO_2 prediction using bi-directional encoder with transformer." Published in Environmental Monitoring and Assessment, Volume 196, Article number 340, (2024)
- **Abhishek Verma,** Virender Ranga, Dinesh Kumar Vishwakarma, "Investigating the Performance vs Computational Complexity Tradeoff in Cross-Domain Fire Risk Detection." Published in Signal, Image, and Video Processing Volume 19, page number 713, (2025)(Pub: Springer)..

Conference Papers

- **Abhishek Verma**, Virender Ranga, Dinesh Kumar Vishwakarma**,** "Forecasting of Satellite Based Carbon-Monoxide Time-Series Data Using a Deep Learning Approach," in 2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT), Institute of Electrical and Electronics Engineers (IEEE), March. 2023, doi: 10.1109/ICITIIT57246.2023.10068609
- **Abhishek Verma**, Virender Ranga, Dinesh Kumar Vishwakarma**,** "Combating Respiratory Health Issues with Intelligent NO_2 Level Prediction from Sentinel 5P Satellite," in 2023 IEEE 20th India Council International Conference (INDICON) at NIT Warangal, Institute of Electrical and Electronics Engineers (IEEE), March. 2023, doi: 10.1109/INDICON59947.2023.10440910

List of Tables

List of Figures

Table of Contents

Chapter 1: INTRODUCTION

The interconnected challenges of wildfires, air pollution, and climate change represent significant global threats with cascading effects on environmental stability, human health, and economic resilience. Wildfires devastate ecosystems and biodiversity and release large amounts of pollutants into the atmosphere, contributing to air quality degradation and the accumulation of greenhouse gases. This increased atmospheric pollution accelerates ozone depletion, exacerbating global warming. The warming atmosphere leads to the melting of polar ice, reducing the Earth's albedo—the reflective capacity of ice surfaces—and causing further heat absorption. This feedback loop accelerates climate change, creating a vicious cycle of environmental degradation. Addressing these interconnected issues requires innovative, efficient, and scalable technological solutions that simultaneously mitigate wildfire risks, predict pollution levels, and monitor climate indicators such as sea ice extent.

The increasing frequency and severity of natural and anthropogenic hazards such as wildfires and air pollution have far-reaching consequences for environmental stability, human health, and economic resilience. Wildfires, in particular, have devastating impacts on ecosystems, leading to biodiversity loss, soil degradation, and atmospheric pollution [1]. For instance, the Australian wildfire catastrophe beginning in September 2019 inflicted over $100 billion in property damage while simultaneously deteriorating soil and air quality and driving multiple species to extinction [2].

Concurrently, air pollution in urban areas, particularly the rising levels of nitrogen dioxide (NO_2), poses a severe threat to public health, contributing to respiratory ailments, cardiovascular diseases, and increased mortality rates[3].

Addressing these challenges necessitates advanced, efficient, and scalable technological solutions. Remote sensing technologies, coupled with machine learning algorithms, have emerged as powerful tools for environmental monitoring, offering high-resolution, real-time data that can enhance predictive capabilities and inform risk mitigation strategies [4], [5], [6].

This thesis explores novel deep learning frameworks for both wildfire risk detection and air pollution prediction, integrating state-of-the-art models and datasets to enhance the accuracy and efficiency of environmental hazard assessments.

1.1 Wildfire Risk Detection

Forests play an indispensable role in maintaining ecological balance, acting as natural carbon sinks, preserving soil integrity, and supporting biodiversity. However, wildfires have increasingly threatened these vital ecosystems, exacerbated by climate change and human activities. The catastrophic wildfires in Australia during 2019-2020 underscore the urgent need for effective fire risk detection and management systems [2].

Traditional fire risk models often rely on geospatial data, satellite imagery, and GIS technologies to map fire-prone zones and predict potential outbreaks [7], [8]. The advent of remote sensing and optical sensor technologies has significantly improved image quality, enabling more precise fire danger evaluations [9]. Recent studies have harnessed machine-learning techniques to analyze remote-sensing images, offering promising results in early wildfire detection [10].

1.2 Air Pollution Prediction

Air pollution remains a pressing issue in urban areas worldwide, with nitrogen dioxide (NO_2) being a major pollutant linked to respiratory and cardiovascular diseases [11]. Delhi, recognized as one of the most polluted cities globally, faces severe air quality challenges due to high population density, industrialization, and vehicular emissions [12].

Satellite-based measurements, particularly from the Sentinel 5P satellite, provide valuable data for monitoring NO_2 concentrations. However, converting satellite-derived tropospheric NO_2 columns into accurate ground-level estimates remains challenging due to atmospheric dispersion and instrument artifacts [13], [14], [15]. This hybrid model effectively captures temporal dependencies and long-range spatial relationships in NO_2 data, significantly improving prediction accuracy [16], [17].

1.3 Sea Ice Classification

Sea ice is a critical component of the polar environment, influencing ocean circulation, climate patterns, and marine ecosystems [18]. Accurate sea ice classification is essential for climate research, marine navigation, and environmental monitoring. Synthetic Aperture Radar (SAR) imagery has been widely used to analyze sea ice dynamics, offering high-resolution, all-weather data crucial for operational monitoring.

Traditional methods for sea ice classification include statistical classifiers, such as the Wishart distribution, and machine learning algorithms, like support vector machines and Markov random fields[19], [20], [21]. However, when trained on small datasets, these methods often struggle with overfitting and limited generalization.

1.4 Motivation

The intricate interplay between environmental hazards such as wildfires, air pollution, and climate change underscores the urgency for innovative monitoring and mitigation strategies. Rising PM2.5 and NO_2 levels, increasing wildfire frequency, and shrinking sea ice pose significant threats to ecosystems, climate stability, and human health. Wildfires degrade air quality and accelerate glacier melt and disrupt atmospheric dynamics, creating cascading effects that further exacerbate global warming. This reduction in sea ice decreases the Earth's albedo, leading to increased heat absorption and accelerating climate change.

To tackle these multifaceted environmental challenges, integrating advanced Artificial Intelligence (AI) and deep learning techniques offers transformative potential. By leveraging satellite data and sensor technologies, AI models can forecast pollution levels, classify sea ice, and assess wildfire risks with unprecedented accuracy and efficiency. The development of systems capable of real-time environmental predictions enhances decision-making processes, enabling proactive responses to emerging threats.

Achieving a balance between computational efficiency and model accuracy is critical for scalable and reliable real-time monitoring. The use of hybrid AI architectures, such as those combining Transformers with BiLSTM models, not only improves prediction accuracy but also ensures robustness and scalability across diverse environmental datasets. These advancements contribute to sustainable solutions, revolutionizing environmental risk prediction, mitigation, and management.

In this context, the motivation for this thesis stems from the pressing need to develop efficient, scalable, and accurate deep learning frameworks that address the interconnected challenges of wildfire risk detection, sea ice classification, and air pollution prediction. By harnessing the power of AI, this research aims to contribute significantly to environmental monitoring, offering innovative solutions for safeguarding ecosystems, public health, and climate stability

1.5 Significance of Study

This study, which focuses on integrating deep learning techniques with remote sensing data for wildfire risk detection, sea ice classification, and air pollution prediction, holds considerable significance for multiple reasons:

- **Enhanced Predictive Accuracy:** This study significantly improves the precision of environmental hazard predictions by combining satellite imagery and advanced machine learning models. Models like Swin Transformer and IGNITE-NET offer superior performance in fire risk detection, while Arctic-Net enhances sea ice classification accuracy, and BREATH-Net improves NO_2 level forecasts in urban settings.
- **Real-Time Environmental Monitoring:** The integration of AI with real-time data sources enables proactive monitoring and timely responses to environmental threats. This capability is crucial for mitigating the immediate impacts of wildfires, pollution, and climate change, providing decision-makers with actionable insights.
- **Contextual Understanding of Environmental Dynamics:** The models developed in this research not only predict events but also provide a deeper understanding of the environmental factors influencing these hazards. This comprehensive analysis helps identify the root causes and interdependencies among wildfires, air pollution, and climate phenomena like ice melt.
- **Robustness to Data Variability:** The deep learning frameworks employed are designed to handle diverse datasets from different geographical regions and environmental conditions. This robustness ensures that the models can be effectively applied in varied real-world scenarios, enhancing their utility and reliability.
- **Applications in Policy and Management:** The findings from this study have direct implications for environmental policy-making and management. By providing accurate predictions and insights, this research supports the development of targeted strategies for wildfire management, air quality control, and climate adaptation.
- **Advancement in AI Applications for Environmental Science:** This thesis demonstrates the potential of cutting-edge AI technologies in solving complex environmental challenges. The novel methodologies proposed contribute to the academic field by expanding the applications of AI in environmental monitoring and management.

In summary, the significance of this study lies in its ability to offer more accurate, real-time, and context-aware solutions for critical environmental issues. By bridging the gap between artificial intelligence and environmental science, this research paves the way for innovative approaches to safeguard natural ecosystems and human health in the face of escalating environmental threats.

1.6 Sources of Research Works Studied

In this thesis, extensive literature reviews and analyses were conducted using leading journals and conference proceedings sourced from the following databases:

- **Elsevier**
- **IEEE Xplore**
- **Springer Link**
- **Association for Computing Machinery (ACM) Digital Library**
- **Taylor & Francis Online**

In addition to these primary sources, more than 150 research papers were screened, focusing on advanced machine-learning techniques, remote sensing applications, and environmental monitoring. Approximately 120 of these papers were selected from high-impact journals and top-tier conferences, ensuring the inclusion of the most recent and influential research.

The search strategy involved utilizing multiple keywords and synonyms, including "wildfire risk detection," "remote sensing for environmental monitoring," "deep learning in climate science," "sea ice classification with SAR imagery," and "air pollution prediction using satellite data." This comprehensive approach ensured a robust foundation for developing the models and methodologies presented in this thesis.

1.7 Overview of Chapters

The remaining sections of this thesis are structured as follows:

- **Chapter 2: Literature Review** – This chapter presents a comprehensive review of existing research and methodologies in environmental monitoring and risk detection. It covers air pollution forecasting, sea ice classification, and wildfire risk detection using advanced deep

learning frameworks. The review integrates findings from key research contributions that have significantly advanced these fields.

- **Chapter 3: PM2.5 Prediction using CATALYST** – This chapter elaborates on the development of the CATALYST model, a hybrid Convolutional Neural Network (CNN) and Transformer-based architecture for predicting PM2.5 concentrations in urban areas. The chapter details the methodology, experimental setup, and performance evaluations, demonstrating the model's superiority over traditional forecasting techniques.
- **Chapter 4: NO_2 Forecasting with BREATH-Net** – This chapter introduces BREATH-Net, a novel deep learning framework that combines Bi-directional Long Short-Term Memory (BiLSTM) networks and Transformer architectures for accurate NO_2 concentration predictions. The chapter discusses data preprocessing, model architecture, and comparative performance analysis with existing models.
- **Chapter 5: Sea Ice Classification using Arctic-Net** – This chapter presents Arctic-Net, a hybrid deep learning model integrating CNNs and attention mechanisms for efficient sea ice classification using SAR images. The chapter covers the architectural components, dataset preprocessing, and performance evaluations, highlighting the model's applicability in climate research and marine navigation.
- **Chapter 6: Wildfire Risk Detection with IGNITE-NET** – This chapter focuses on IGNITE-NET, an innovative deep-learning framework designed for wildfire risk prediction. It explores the use of Dynamic Receptive Field Blocks (DRFBs) and Dynamic Channel Fusion Attention (DCFA) to enhance predictive accuracy while maintaining computational efficiency. The chapter includes detailed performance evaluations and comparative analyses.
- **Chapter 7: Performance vs Computational Complexity in Fire Risk Detection** – This chapter investigates the trade-offs between model performance and computational complexity in cross-domain fire risk detection, emphasizing the Swin Transformer architecture. It provides a thorough exploration of methodologies, experimental setups, and performance metrics, along with future research directions.
- **Chapter 8: Conclusion, Future Scope, and Social Impact** – This concluding chapter summarizes the key contributions of the thesis, discusses the broader implications of the research findings, and outlines potential directions for future work. It also highlights the

social and environmental impact of the developed models in addressing global challenges related to wildfire management, climate monitoring, and urban air quality control.

This chapter has introduced the pressing environmental challenges posed by wildfires, air pollution, and climate change, highlighting the need for advanced technological solutions.

Building upon these issues, the next chapter provides a comprehensive review of existing research and methodologies that address these challenges, particularly focusing on the use of deep learning frameworks for air pollution forecasting, wildfire risk detection, and sea ice classification. This review forms the foundation for understanding the state-of-the-art approaches that are critical for developing effective solutions.

Chapter 2: LITERATURE REVIEW

This chapter comprehensively reviews existing research and methodologies in environmental monitoring and risk detection. It focuses on air pollution forecasting, sea ice classification, and fire risk detection using advanced deep learning frameworks. The review integrates findings from key research contributions that have significantly advanced these fields. This chapter presents a comprehensive review of existing research and methodologies in environmental monitoring and risk detection. It focuses on air pollution forecasting, sea ice classification, and fire risk detection using advanced deep learning frameworks. The review integrates findings from key research contributions that have significantly advanced these fields.

2.1 Air Pollution Forecasting and Analysis

Air pollution is a significant environmental and public health issue, with PM2.5 and NO_2 being among the most harmful pollutants. Traditional forecasting models, such as ARIMA and SVR, have shown limitations in capturing the complex, non-linear nature of pollution data. Recent advancements in machine learning and deep learning have significantly improved the accuracy of air quality predictions.

2.1.1 Advanced Machine Learning and Deep Learning Techniques for PM2.5 Concentration Prediction

Air pollution is a significant problem affecting millions of people's quality of life globally. The World Health Organization (WHO) estimates that air pollution causes 7 million premature deaths each year, with particulate matter (PM2.5) being one of the most harmful pollutants[22]. The investigation concerning historical data emphasizes the importance of comprehending the patterns of air pollution over a span of time, providing valuable insights into the progression of air quality in densely populated urban areas [23]. Predicting PM2.5 concentrations accurately is critical for mitigating air pollution's adverse effects. Researchers have been exploring various machine learning models to estimate PM2.5 concentrations accurately. In recent years, Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have been the most popular models used[24], [25], [26], [27].

This literature review aims to provide an overview of the recent research on predicting PM2.5 concentrations using machine learning models. CNNs, extensively employed in image processing, have been utilized in studies to forecast air quality. Liu et al.[28] outperformed standard methods such as ARIMA Jian et al.[29] by utilizing a pre-trained CNN model to anticipate PM2.5 concentrations using meteorological data. Similarly, Yao et al. [26] employed a CNN-based model with multi-scale and multi-channel characteristics. Li et al. [30] extracted spatial data from PM2.5 concentration maps and captured temporal features using an LSTM network[30], [31]. Both models outperformed traditional approaches such as SVR and ANN [32].RNNs, which can model sequential data, have been used to forecast time series[33] and predicted PM2.5 concentrations using a hybrid model that combined a GRU and a CNN (Zhang, 2020), surpassing standard models such as ARIMA and SVM. The study by Shi et al. [34] utilized meteorological data to train an attention-based Recurrent Neural Network (RNN) model. The study's results indicated that the attention-based RNN model's performance surpassed conventional methods such as Autoregressive Integrated Moving Average (ARIMA) and Random Forest (RF). Hybrid models have also been proposed, combining machine-learning techniques with classical statistical methodologies. Liu et al. [25] created a hybrid model combining a CNN and a seasonal decomposition approach, outperforming classic methods like ARIMA and SVM[25]. Wang et al.[35] suggested a hybrid model that beat established procedures such as ARIMA and SVM by combining a wavelet transform with a multivariate adaptive regression splines (MARS) approach [36]. Incorporating additional data sources, such as meteorological and traffic data, has been demonstrated to enhance PM2.5 forecasting accuracy. Feng et al.[37] meteorological and traffic data were used to estimate PM2.5 concentrations in the Beijing–Tianjin–Hebei region, outperforming established methods such as ARIMA and SVM. Wang et al.[38] projected PM2.5 concentrations using meteorological data and a MARS technique, surpassing established methods such as ARIMA and SVM. Anthropogenic factors, air quality, and PM2.5 concentrations are also influenced by anthropogenic causes such as industrial growth and transportation. The study by Zhu et al. [39] aimed to examine the influence of industrial emissions on the concentrations of PM2.5 in China. The study's findings revealed that the high levels of PM2.5 in the country were primarily attributed to the emissions from the industrial sector. Similarly, transportation emissions from vehicles such as cars and trucks are also a significant source of air pollution, particularly in urban areas with high traffic congestion[40]. In addition, changes in land use, such as

deforestation, urbanization, and agricultural practices, can also contribute to air pollution. Deforestation can potentially elevate the levels of dust and smoke in the atmosphere, whereas urbanization may result in amplified emissions from industrial and transportation activities[41]. Ultimately, climate change can exert a substantial influence on the quality of the air. Elevated temperatures have the potential to augment the occurrence and severity of wildfires, resulting in escalated concentrations of smoke and particulate matter within the atmosphere. The phenomenon of climate change has the potential to cause alterations in precipitation patterns, thereby influencing the dispersion of pollutants in the atmosphere and their conveyance to diverse geographical locations.

2.1.2 Advanced Machine Learning and Deep Learning Techniques for NO_2 Concentration Forecasting

The growing focus on the impacts of air pollution on human health and the environment is highlighted by recent improvements in air pollution forecasting. This literature review consolidates findings from various research utilizing novel methodologies and frameworks, enhancing the continuous endeavors in environmental management and safeguarding public health. Heydari et al. [42] proposed a hybrid intelligent model for predicting air pollution. The model combines a long short-term memory (LSTM) deep learning model with the multi-verse optimization algorithm (MVO)[42]. The suggested model demonstrated more accuracy in predicting Nitrogen dioxide and sulfur dioxide (SO2) emissions from Combined Cycle Power Plants compared to benchmark models (ENN-PSO, ENN-MVO, LSTM-PSO). The study's focus on accurate forecasts underscores the changing nature of forecasting models in dealing with the intricacies of air pollution dynamics. Huang et al.[43] examined the issue of air pollution in China, with specific emphasis on exposure to NO_2. Their integrated model, which combines satellite measurements, simulations from a chemical transport model, and detailed geographical factors, exhibited exceptional accuracy in forecasting daily NO_2 values between 2013 and 2019[44]. This work made significant contributions to exposure modeling and also uncovered a declining pattern in NO_2 exposure, offering vital insights into the regional and temporal dynamics of air quality in China. An extensive modeling framework was created to anticipate air pollution in Jiangsu Province, China. Douros et al.[45] integrate satellite measurements, simulations from a chemical transport model, and machine learning techniques. The ensemble model demonstrated effectiveness in

forecasting daily NO_2 levels over a significant period, highlighting the need for reliable predictor data in achieving precise air pollution predictions.

Rakholia et al. [46] have suggested a new and complex model for predicting air quality in Ho Chi Minh City, Vietnam. This model has many steps, outputs, and variables. The global forecasting model incorporates multiple parameters, including meteorological conditions, air quality data, urban space information, and time components. It has proven its ability to simultaneously predict various air pollutant concentrations, surpassing previous forecasting methods. Using artificial intelligence, Hasnain et al.[47]employed a novel way to monitor ground-level NO_2 levels throughout China. By incorporating spatiotemporally weighted information into extra trees and deep forest models, the research successfully addressed the absence of satellite data, creating a detailed dataset called "ChinaHighNO_2." This dataset enables detailed analysis of spatial patterns and the effects of holidays and the COVID-19 pandemic. Wei et al. [48] presented a unique technique for estimating the emissions and lifetimes of nitrogen oxides (NOx) in cities. This method relies on observing tropospheric NO_2 levels and analyzing wind patterns using reanalysis data. The study demonstrated the precision of the technique in calculating emissions and lifespans for 26 cities in the United States, highlighting its capacity to estimate worldwide urban NOx emissions from satellite measurements. An extensive analysis highlighted the crucial importance of artificial intelligence (AI) techniques and machine learning (ML) algorithms in predicting air pollution and its effects on human health [48]. The evaluation emphasized the effectiveness of hybrid models in accurately and reliably forecasting different significant pollutants, highlighting their superiority over individual AI models.

The research on time series forecasting of air quality in Sofia City, Bulgaria, provided valuable insights into air pollution patterns using the Auto Regressive Integrated Moving Average (ARIMA) technique [49]. By examining data collected between 2015 and 2019, the study has deepened our comprehension of how pollutants behave over time. This knowledge has played a crucial role in informing measures to prevent and regulate pollution, ultimately leading to higher air quality. Guo and Mao [50] have contributed noteworthy to developing air quality forecasting models by introducing a unique long-term prediction model designed explicitly for NOx emissions. By integrating self-attention to capture long-term patterns and utilizing a parallel LSTM-Transformer architecture, their solution exhibits significant enhancements, outperforming

existing techniques by 28.2% and 19.1% on separate datasets. The research shows the potential for accurate regulation of NOx emissions in response to strict environmental regulations. To summarise, the literature indicates a significant increase in advanced modeling approaches that utilize sophisticated technology for predicting air pollution. Combining hybrid intelligence models, high-resolution exposure modeling, and unique city-specific emission inference methods, all contribute to current endeavors in environmental management and safeguarding public health.

2.2 Innovations in Sea Ice Classification Using Deep Learning

Sea ice has a crucial impact on the global climate system, especially in the Polar Regions, since it affects ocean circulation, regional weather patterns, and the Earth's albedo. Understanding the creation and dynamics of sea ice is crucial because of its profound influence on marine ecosystems, nutrient cycling, and global climate systems. Advancements in remote sensing, particularly in SAR imaging, have allowed in-depth investigations of sea ice dynamics. This literature review focuses on the progression of methodology for sea ice categorization using SAR data, particularly emphasizing the shift from conventional statistical methods to more sophisticated machine learning techniques. Recent progress in categorizing sea ice using SAR images has brought out novel approaches to improve precision and effectiveness. Various technologies, including local thresholding techniques and advanced deep learning frameworks, aim to overcome the limits of classic methods and offer novel perspectives on sea ice monitoring.

Researchers comprehensively explained a dynamic local thresholding approach that adjusts to the local fluctuations in grey levels inside SAR pictures. This technique enhances the accuracy of differentiating ice types by dynamically adapting to regional changes in the picture, surpassing previous global thresholding methods. The study's results suggest that this approach, which does not require any user involvement, is relatively successful at separating ice initially. It may be improved even more by utilizing expert systems to categorize particular ice forms and establish their proportions within the images [51].

Notable progress has been made by utilizing a sophisticated deep neural network, MSI-ResNet, to accurately categorize different sea ice forms during the late spring and summer seasons, utilizing GF-3 quad-polarization data. This study highlights the significance of selecting the most suitable patch sizes and polarization combinations to achieve a high level of accuracy in categorization.

The MSI-ResNet approach performed more excellently than conventional classifiers, such as SVM, with high overall accuracy and kappa coefficients across several Arctic areas. Deep learning in this study dramatically enhances the accuracy of sea ice type recognition [52]. A new advanced deep learning framework called Deep SAR-Net has also been created. This framework is specifically built to handle complex-valued SAR pictures. Deep SAR-Net effectively captures spatial characteristics and backscattering patterns by combining intensity pictures with radar spectrograms, improving accuracy in discriminating objects. This approach performs superior to traditional deep convolutional neural networks (CNNs), particularly in differentiating objects with similar textures but in their scattering properties. It does this by using both spatial and frequency domain characteristics [53].

Furthermore, a hierarchical deep learning-based pipeline mapping sea ice from SAR pictures has been suggested. This method utilizes a semantic segmentation model to map the boundaries between ice and water accurately. Then, it applies a two-level hierarchical CNN to classify the ice in finer detail. The hierarchical strategy enhances classification accuracy by explicitly addressing the unequal visual differentiation of various ice kinds, surpassing the performance of typical flat N-way classifiers [54]. A more sophisticated technique for classifying sea ice has been created, which integrates polarization information obtained from polarization decomposition with spectrogram features derived from joint time-frequency analysis (JTFA). This method provides good classification accuracy using ALOS PALSAR SLC data with quad-polarization. It requires less data and computing effort compared to single-feature methods. The study showcases the ability to maximize the potential of SAR data by integrating several characteristics, resulting in a substantial improvement in classification accuracy [55].

The research introduces a unique technique called Physically Explainable CNN for SAR image classification called physics-guided and injected learning (PGIL). PGIL incorporates the distinct electromagnetic properties of SAR data into the deep learning framework to improve comprehensibility and understanding of physics. This approach consists of three components: explainable models (XM) that offer previous knowledge of physics, a physics-guided network (PGN) that encodes this information into physics-aware features, and a physics-injected network (PIN) that incorporates these features into the classification process. The assessments conducted using Sentinel-1 and Gaofen-3 SAR data show that PGIL significantly enhances classification

accuracy, especially when labeled data are scarce, compared to conventional CNNs and pre-training techniques. The work emphasizes the capacity of PGIL to retain physical coherence in predictions, avoid overfitting, and uphold interpretability through physics-guided signals, providing a robust and comprehensible method for SAR image categorization[56].

These studies demonstrate substantial advancements in sea ice categorization using SAR technology. Researchers have utilized dynamic local thresholding, sophisticated deep learning frameworks, and multi-feature techniques to enhance the precision and dependability of methods used for monitoring and analyzing sea ice. These developments improve our comprehension of sea ice movements and offer essential instruments for monitoring the environment and studying climate. Further advancement and fine-tuning of these techniques will enhance the accuracy and suitability of SAR-based sea ice categorization.

2.3 Advanced AI Approaches for Wildfire Risk Prediction and Management

Wildfires pose significant threats to ecosystems, human life, and infrastructure. Accurate fire risk prediction is crucial for proactive wildfire management and disaster response. Machine learning and deep learning approaches have been increasingly utilized to enhance the accuracy and efficiency of fire risk detection systems.

2.3.1 Dynamic Approaches to Fire Risk Prediction

In recent years, fire risk detection has become a crucial area of research due to the growing threat of wildfires to ecosystems, economies, and human life. Many methods have been developed that center on striking a compromise between detection performance and computational complexity. It is evident from Munsif et al. [57]. Convolutional Neural Networks (CNNs) are widely utilized for fire detection, and a lightweight model was suggested for disaster recognition. This efficient CNN model was implemented on devices with limited resources, proving its usefulness in practical situations.

Similarly, IoT-based fire detection systems have been developed, using CNNs to interpret data in real time. With an emphasis on Internet of Things contexts, Dilshad et al. [58]suggested an optimized fire attention network (OFAN) for efficient fire detection. By addressing issues with accuracy in dimly lit and foggy environments, their approach produced better results across several

datasets. The model's ability to capture global context was enhanced by adding a dilated convolutional layer, which qualified it for real-time applications on edge devices. Moreover, recent developments in vision transformers (ViTs) have demonstrated promise for fire detection. However, ViTs' high computational demand prevents them from being used in settings with limited resources. To address these issues, a unique ViT architecture that combines local self-attention with shifting patch tokenization was presented by Yar et al. [59]. This method made effective fire detection possible even with small and medium-sized datasets. The changes to the model balanced computational cost and accuracy by reducing the model size and the number of floating-point operations. To improve the accuracy of fire detection, several designs have added attention methods in addition to CNN and ViT models. For example, Yar et al. [60] provided a model that combines 3D convolution operations with a modified soft attention mechanism to enhance the identification of tiny fires from UAV data. Their method clarified how crucial it is to control model complexity to facilitate real-time UAV deployment, which is essential for prompt action in regions where fires are likely to occur.

Furthermore, models based on attention are still developing. The multi-attention fire network (MAFire-Net), introduced by Khan et al. [61], incorporates channel and spatial attention processes to improve feature representation in fire detection tasks. Compared to state-of-the-art techniques, this architecture demonstrated greater accuracy and faster inference times, which makes it a good contender for real-time deployment on edge devices. Its improved performance over several benchmarks was further enhanced by creating a sizable fire dataset. The advancement of fire detection technology has also been significantly aided by dataset creation. A substantial source of high-resolution UAV-captured photos to detect forest fires is the UAVs-FFDB dataset, first presented in UAVs-FFDB [62]. This dataset contributes to creating more resilient AI models for fire detection, monitoring, and diversifying the training set. Researchers may develop and test new fire detection algorithms using UAV-collected data, which enhances real-time fire monitoring systems.

Additionally, ensemble approaches have been suggested to increase the accuracy and resilience of fire detection models. Belarbi et al. [63] created a CNN ensemble to categorize fires in their early stages, providing a low-cost substitute for deep CNNs, which are frequently computationally

costly. This method is perfect for early fire warning systems since it aggregated predictions from several models, obtaining great accuracy at a low computing cost.

Ultimately, multi-scale techniques have effectively addressed the difficulties associated with early fire detection, mainly when dealing with overlapping fire targets and challenging-to-detect smoke pictures. Introduced by Yan et al. [64], the multi-scale depth-separable convolutional network (MDCNet) was created to capture complex fire properties of several sizes. The application of focus loss improved its capacity for difficult fire situations much further. With its ability to identify fires earlier than conventional fire detection techniques, this design showed promise for improving public safety.

The research described in these publications offers a range of methods that address performance, computing efficiency, and practical application, which substantially contribute to the continuous development of fire detection systems. When taken as a whole, these studies offer a thorough framework for improving fire risk assessment, especially when it comes to cross-domain fire detection, which is the focus of the current study

2.4 Research Gaps

- Limited integration of advanced hybrid models for addressing spatiotemporal complexities in environmental data.
- Insufficient utilization of satellite data combined with ground monitoring for air quality forecasting.
- Minimal exploration of computationally efficient models for large-scale SAR image classification with high accuracy.
- Inadequate research on balancing performance and computational complexity in fire risk detection models.
- Limited application of self-supervised learning techniques like knowledge distillation in environmental risk prediction.
- Few studies focus on real-time deployment and scalability of models for practical scenarios across diverse regions.

2.5 Research Objectives

The following objectives have been proposed based on the identified research gaps:

- To develop AI-based solutions to predict accurate pollution levels for efficient policymaking.
- To propose an efficient method for detecting forest fire risk using hyperspectral satellite imagery.
- To design a novel AI-driven framework for identifying sea ice in Synthetic Aperture Radar (SAR) data.

This chapter provides an extensive review of current research on environmental monitoring, highlighting significant advancements in air pollution forecasting, sea ice classification, and fire risk detection through deep learning techniques. It explores the limitations of traditional methods and sets the foundation for more advanced solutions. The next chapter will build upon these findings by introducing a novel hybrid deep learning model that addresses the complexities of PM2.5 forecasting, demonstrating how the integration of convolutional and transformer-based architectures can significantly improve predictive accuracy and computational efficiency.

Chapter 3: CATALYST: A NEW ARCHITECTURE FOR PREDICTION of PM 2.5

3.1 Scope of this Chapter

In the modern era of rapid industrialization and urbanization, air pollution has become a pressing global concern, significantly affecting human health and environmental sustainability. Among various pollutants, PM2.5 (Particulate Matter with a diameter of 2.5 micrometers or less) poses severe health risks due to its ability to penetrate deep into the human respiratory system, leading to respiratory diseases, cardiovascular complications, and reduced life expectancy. The increasing concentration of PM2.5 in metropolitan cities like Delhi has prompted the need for accurate forecasting models that can help mitigate its adverse effects and support evidence-based policymaking. Conventional forecasting techniques such as statistical models (ARIMA) and machine learning approaches (SVR, Decision Trees, and Random Forests) have demonstrated limited success in accurately predicting PM2.5 concentrations due to their inability to fully capture the non-linear and highly dynamic nature of air pollution data. These methods struggle to integrate multiple influencing factors, such as meteorological conditions, vehicular emissions, industrial activities, and seasonal variations, into a cohesive prediction model. To address these limitations, this research introduces CATALYST, a hybrid Convolutional Neural Network (CNN) and Transformer-based model designed to enhance the accuracy of PM2.5 forecasting. The CATALYST model effectively leverages deep learning techniques to integrate spatial and temporal dependencies in air pollution data, providing more reliable and robust predictions. By combining CNNs for feature extraction and Transformers for sequential learning, CATALYST aims to improve forecasting precision, optimize computational efficiency, and outperform traditional prediction models. This chapter delves into the development, implementation, and evaluation of CATALYST for PM2.5 forecasting in Delhi, demonstrating its superiority over conventional models. The proposed model aims to enhance real-time air quality monitoring, facilitate proactive pollution control strategies, and contribute to global efforts in environmental sustainability.

3.2 A Novel Approach for Forecasting PM2.5 Pollution in Delhi Using CATALYST

3.2.1 Abstract

Air pollution, particularly PM2.5, poses significant threats to public health and environmental stability, necessitating robust predictive models for effective mitigation strategies. Traditional forecasting approaches, including statistical models (e.g., ARIMA) and machine learning techniques (e.g., SVR, Decision Trees, and LSTM), have shown limitations in capturing the complex spatial-temporal dependencies of PM2.5 pollution. To overcome these challenges, this study proposes CATALYST, an advanced hybrid deep learning model integrating Convolutional Neural Networks (CNNs) and Transformer architectures to enhance prediction accuracy. The CNN module efficiently extracts spatial features from air pollution datasets, while the Transformer component leverages self-attention mechanisms to model long-term temporal dependencies. Extensive experiments were conducted using 48,362 hourly PM2.5 records from five monitoring stations in Delhi, comparing CATALYST against state-of-the-art forecasting models such as ARIMA, LSTM, and standard Transformer-based approaches. The results demonstrate that CATALYST achieves the lowest RMSE (21.01) and the highest R^2 (0.89), outperforming all baselines in predictive accuracy. Furthermore, CATALYST exhibits superior generalization capabilities, making it adaptable to different air quality datasets. This research contributes to advancing AI-driven environmental monitoring by offering an innovative deep-learning framework for air quality forecasting. The findings of this study provide valuable insights for policymakers and urban planners in designing data-driven pollution control strategies, thus supporting global sustainability efforts. Air pollution, particularly PM2.5, significantly impacts human health and environmental quality. Predicting PM2.5 concentrations accurately is crucial for effective air quality management. Traditional statistical and machine learning models have struggled to provide robust forecasts due to the complex spatiotemporal nature of air pollution. This study introduces CATALYST, a hybrid CNN-Transformer deep learning model, designed to improve the accuracy of PM2.5 forecasting in Delhi. The proposed model effectively captures both short-term and long-term dependencies in air pollution data, outperforming existing statistical and deep learning approaches. The experimental results demonstrate that CATALYST achieves superior performance compared to ARIMA, LSTM, and Boosting models, with lower RMSE and

higher R^2 values. This research contributes to advancing air pollution forecasting and aiding policymakers in implementing effective mitigation strategies.

3.2.2 Proposed Methodology

This section presents a comprehensive overview of the proposed methodology, which is divided into three main phases. The first phase involves the collection and pre-processing of data obtained from five distinct monitoring stations located in Delhi: DTU Delhi-CPCB, NSUT Dwarka-CPCB, Anand Vihar-DPCC, Pusa Delhi-IMD, and IGI-T3 Delhi-IMD. An overall model of the methodology is depicted in Figure 3.1. Subsequently, the gathered information is processed to facilitate subsequent examination. The following phase involves the utilization of a pioneering Transformer-based methodology to predict air quality. The data that has been gathered and undergone pre-processing from the monitoring stations are utilized as input for the Transformer model. By using its attention mechanism, the Transformer model can effectively capture extended dependencies and acquire intricate patterns within the data, ultimately facilitating precise predictions regarding air quality.

During the third stage, the outcomes derived from the predictive model are evaluated and scrutinized through the application of performance metrics, including but not limited to mean absolute error (MAE), coefficient of determination (R^2), and root mean square error (RMSE). The present study examines the precision and efficacy of the suggested approach in forecasting atmospheric conditions. The proposed methodology integrates the various components such as data collection, pre-processing, a unique Transformer-based approach, and result analysis to offer comprehensive framework for air quality prediction.

Table 3.1 shows the algorithm of the proposed architecture. The algorithm's objective is to facilitate the training and prediction of air quality by utilizing the CATALYST model. The system receives input in the form of data points obtained from monitoring stations and uses this information to forecast air quality values. The iterative process of the algorithm is executed for a predetermined number of epochs. The process involves converting the input data into a vector image and extracting features through a Convolutional Neural Network (CNN). The Transformer block is used to extract enduring characteristics. The application of dropout is observed in the convolutional neural network features, while the Transformer features are subjected to both

dropout and batch normalization. The final predicted output is derived by concatenating the enhanced features.

Table 3.1 Training and prediction algorithm for **CATALYST**

Aim: To learn a mapping function $\mathbb{F}: (\boldsymbol{X_j}, \boldsymbol{Y_j}) \longrightarrow$ from data points obtained from five monitoring stations

Input: Set of data points $\boldsymbol{X_j} = \{\boldsymbol{x_{j1}}, \boldsymbol{x_{j2}}, \ldots\ldots\ldots, \boldsymbol{x_{jSW}}\}$ where, $\boldsymbol{SW}$ is the size of the sliding window used for input layer partitioning

Output: Air Quality Prediction, $\boldsymbol{Y_j} \in \boldsymbol{forcasted\ value}$

1. $D = \{(\boldsymbol{X_0}, \boldsymbol{Y_0}), (\boldsymbol{X_1}, \boldsymbol{Y_1}), \ldots\ldots\ldots\ldots, (\boldsymbol{X_B}, \boldsymbol{Y_B})\}$ represents the dataset consisting of hourly data points $\boldsymbol{X}$ and $\boldsymbol{Y}$ as the forecasted value with $\boldsymbol{B}$ number of blocks into which the data is partitioned for both input and label data.

   for $\boldsymbol{E} \leftarrow 1$ to *Epochs* do

2. Input data $\boldsymbol{X_j}$ is converted into a vector image.
3. $\mathcal{F}(\boldsymbol{X_j})$ be the feature representation obtained by CNN from vector image.
4. $T\left(\mathcal{F}(\boldsymbol{X_j})\right) \leftarrow \mathcal{F}(\boldsymbol{X_j})$ be the long-term features obtained from Transformer block.
5. $D\left(\mathcal{F}(\boldsymbol{X_j})\right) \leftarrow \mathcal{F}(\boldsymbol{X_j})$ improve model's performance by applying dropout to feature representation obtained by CNN.
6. $B\left(D\left(\mathcal{F}(\boldsymbol{X_j})\right)\right) \leftarrow T\left(\mathcal{F}(\boldsymbol{X_j})\right)$ improve model's performance by applying dropout and batch-normalization to feature representation obtained by transformer.
7. $P(\boldsymbol{X_j}) = D\left(\mathcal{F}(\boldsymbol{X_j})\right) \oplus B\left(D\left(\mathcal{F}(\boldsymbol{X_j})\right)\right)$ concatenation of both the enhanced features to obtain forecasted output.

   end

Data Pre-processing

This research work employs a dataset consisting of 48,362 hourly data points that have undergone a process of cleaning and organization. The above data were obtained from five monitoring stations and merged into a unified column. To facilitate analysis, the data underwent a process of partitioning into blocks encompassing input and label data. Utilizing the sliding window technique in the input layer boosted the streamlining of data processing by decreasing nested loops and consolidating them into a singular circle. This methodology conserves both time and computational resources. Furthermore, using the positional encoding layer facilitated the integration of sinusoidal positional embedding with the input data. The process involves converting the PM2.5 data into a vector image, which is subsequently utilized as input for a pre-existing Convolutional Neural Network (CNN) to extract features**.**

Short-Term Contextual Feature Learning using CNN.

So, in the proposed methodology, input is fed as an image to a pre-trained Convolutional Neural Network (CNN). That is employed for feature extraction from the input time series before their integration into the Transformer. Utilizing a pre-existing Convolutional Neural Network (CNN) for feature extraction from PM2.5 data is a viable methodology because CNNs are tailored toward image processing tasks and can proficiently capture spatial patterns and dependencies within the data. By converting PM2.5 data into a vector image, the convolutional neural network (CNN) can extract pertinent features from the data while decreasing its dimensionality. This

facilitates the subsequent transformer model in carrying out temporal modelling with greater efficiency.

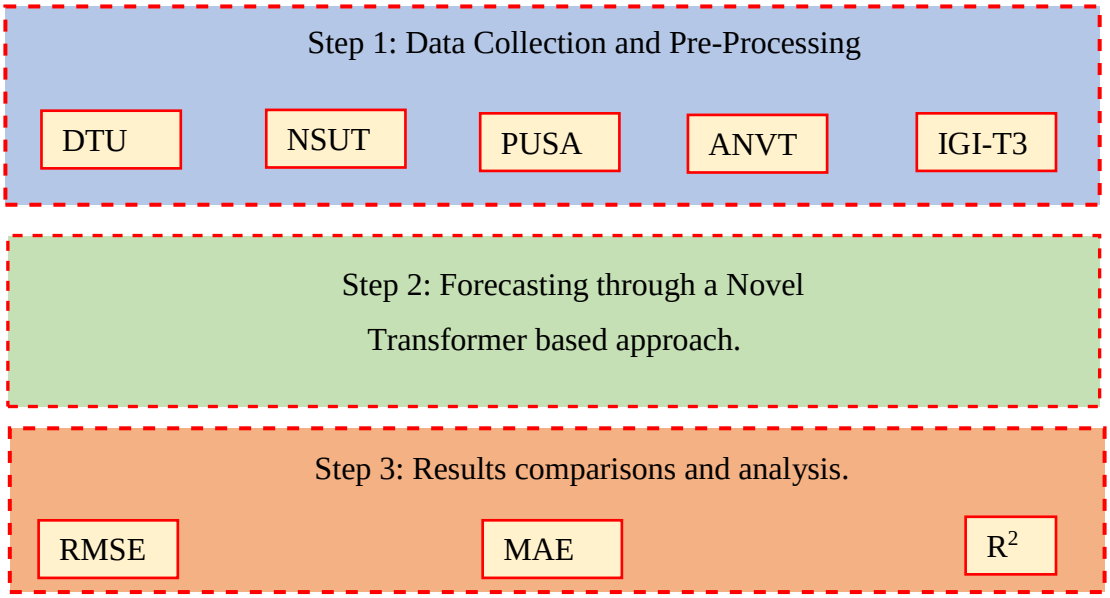


Fig. 3.1 Overall Pipeline of the Proposed Solution

Moreover, utilizing a pre-existing model conserves time and computational resources, since the model has already undergone comprehensive training on vast datasets. It can be adapted to the particular task at hand through fine-tuning. Finally, using the sliding window technique in the input layer diminishes the occurrence of nested loops and streamlines the temporal intricacy of the model. In general, the rationales, as mentioned earlier, render the utilization of a pre-trained convolutional neural network a viable methodology for predicting PM2.5 data.

Long-Term Contextual Features Using a Transformer

Long-term contextual characteristics are just as crucial for the effective forecasting of air quality data as the short-term contextual variables provided by the pre-trained CNN. The suggested solution uses a transformer network to capture these long-term relationships. The transformer network may model the input data sequence's long-term relationships well, which can also capture the context necessary for precise prediction. The primary task of the transformer encoder layers involves using self-attention and feed-forward neural network operations to carry out the prognosis. Ultimately, the decoder's linear layer produces conclusive prognostications. Full graphical representation of architecture model can be seen in Fig. 3.2.

The transformer network computes a sequence of outputs that may be utilized for predicting based on a series of inputs. It comprises several encoder layers, each with a feed-forward neural network and a self-attention mechanism. The feed-forward neural network gives the model non-linearity, and the self-attention tool enables the transformer to recognize the significance of each input piece.

The mean squared error (MSE) loss function and Adam optimizer used in CNN training are also used in the transformer. During training, a validation set is used to keep track of the model's progress while a batch size of 64 is set. A cyclical learning rate schedule is utilized to improve convergence, and the learning rate is adjusted during training using the learning rate annealing approach.

In general, integrating a pre-trained Convolutional Neural Network (CNN) and a Transformer network facilitates the extraction of contextual features from the input time series, encompassing both short-term and long-term information. This enables precise prediction of atmospheric quality information with reduced computational resources.

Final prediction

The ultimate forecast is generated through the amalgamation of the pre-existing Convolutional Neural Network and the Transformer network's outputs. The convolutional neural network that has been pre-trained can extract contextual features that are short-term in nature. On the other hand, the Transformer is used to capture the long-term relationships in the input data. The final prediction is derived by integrating the outputs of both models.

To enhance the model's performance, several techniques, including dropout and batch normalization, are employed to mitigate the issue of overfitting and augment the model's capacity to generalize the novel data. The evaluation of the model's performance is conducted through the utilization of diverse evaluation metrics, including but not limited to Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2).

In brief, the ultimate forecast is derived by amalgamating the results obtained from the pre-existing Convolutional Neural Network (CNN) and the Transformer network, followed by implementation methodologies such as dropout and batch normalization to enhance the efficacy

of the model. The evaluation of the model's performance is conducted through diverse evaluation metrics. The complete methodology is presented in

Fig. 3.1.

Computational efficiency

Various techniques are investigated and assessed with the aim of improving computational efficacy and mitigating the risk of overfitting. The selected model utilizes a solitary encoder and a linear layer, resulting in a notable decrease in complexity while still achieving proficient performance.

The utilization of a solitary encoder in the model results in optimization of the processing pipeline, thereby, reducing the computational burden. The utilization of this methodology obviates the necessity for numerous encoder layers, thereby, diminishing the computational workload during both the training and inference stages. Furthermore, the model integrates a linear layer, which enhances the computational efficiency. Linear layers exhibit computational efficiency in comparison to more intricate layers, such as fully connected layers, due to their utilization of simpler matrix operations. Through the implementation of these measures, the model can enhance the efficiency of computational resources while maintaining the precision of its predictions. The acceleration of training and inference times facilitates the model's practicality for real-time or large-scale applications. In general, the adoption of a solitary encoder and a linear layer amplifies computational efficacy, thereby, expediting the processing and mitigating the likelihood of

overfitting. The enhancements facilitate a heightened level of efficacy and pragmatic in the forecasting framework.

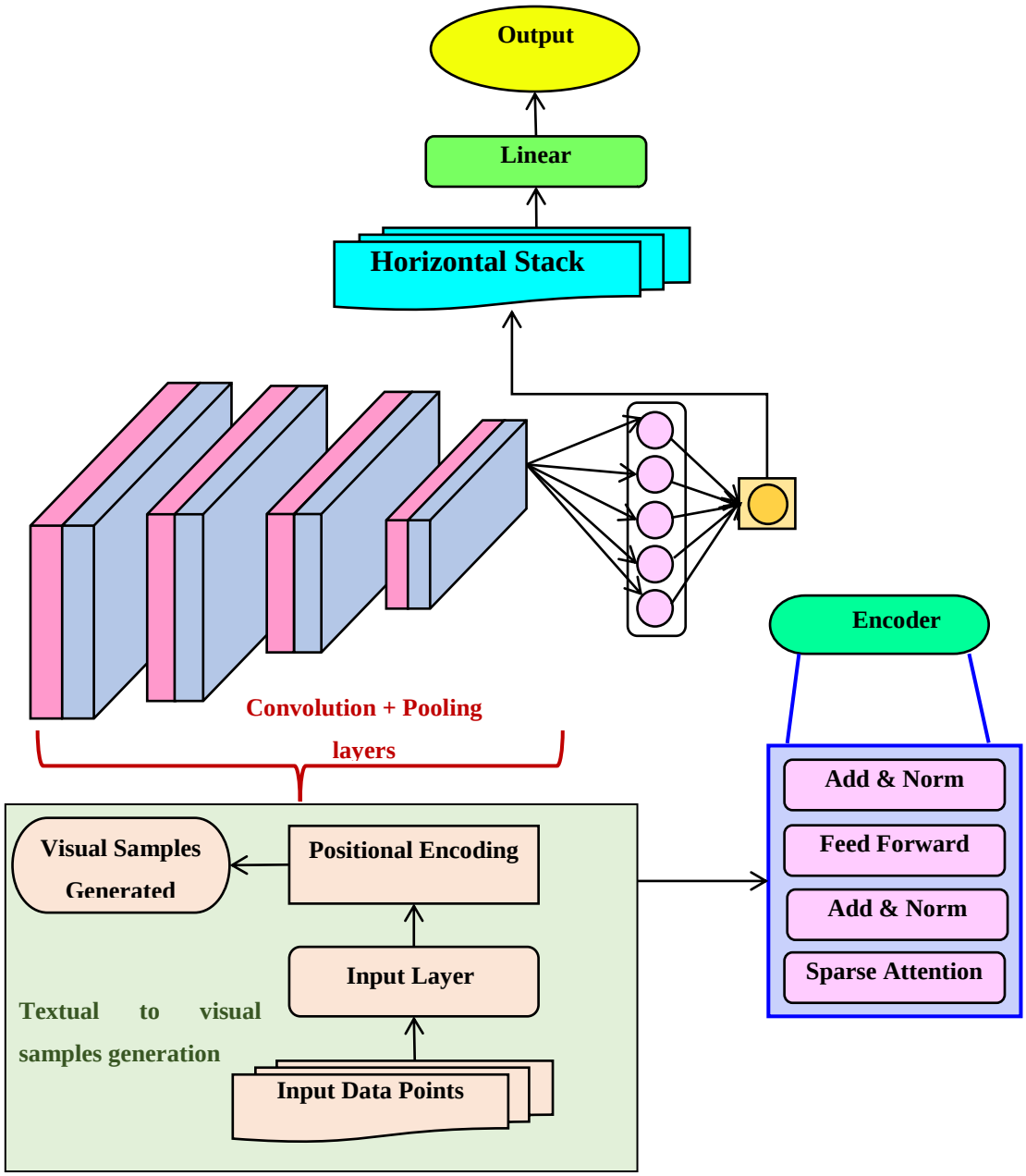


Fig. 3.2 The architecture of CATALYST -a novel Convolutional and Transformer model for Air Quality Forecasting

This section outlines the technical aspects of the study's implementation, encompassing the hardware configuration, employment of machine learning frameworks, and the particular model architecture utilized. The present discourse delves into the process of optimizing the model, the parameters involved in training, and the metrics employed for evaluation. Furthermore, the section

presents a comprehensive examination of the dataset, demonstrating patterns in PM2.5 levels across various temporal dimensions such as time, seasons, and hours. The discourse underscores the association between the concentration of PM2.5 and diverse factors, including but not limited to the day of the week, season, and atmospheric conditions. Additionally, the aforementioned section conducts a comparative analysis between the CATALYST model and other baseline models, showcasing its heightened predictive accuracy and dependability.

Implementation Details

The experiments are conducted on a PC server equipped with an NVIDIA QUADRO RTX A5000 graphics card featuring a memory capacity of 24GB and 3042 NVIDIA Cuda cores. Implementing machine learning techniques for air quality forecasting involves the utilization of open-source frameworks such as Pytorch (https://pytorch.org) and Sklearn (https://scikit-learn.org/). The open-source Tensorflow library, available at https://github.com/tensorflow/, configures deep learning and Transformer models. The implementation that has been presented is founded on a time series forecasting model that utilizes transformers. This model combines a convolutional neural network (CNN) and a transformer network.

The model optimization process involves utilizing the mean squared error (MSE) loss function alongside the Adam optimizer for updating the model parameters. The model's training is conducted with a batch size of 64. The training process is terminated after 1000 epochs or when the validation loss fails to demonstrate improvement for ten consecutive periods. The code employs an automatic technique, learning rate annealing, to modify the learning rate in response to the training progress. Additionally, cyclical learning rate schedules vary the learning rate to enhance convergence cyclically. The starting learning rate varies from 0.000001 to 0.000200 according to requirement. The proposed framework's training, validation, and testing ratio has been set to 80%, 10%, and 10%, respectively as it can be seen in Fig. 3.3. The optimal model is selected employing the validation loss metric and subsequently employed to generate predictions on the test dataset.

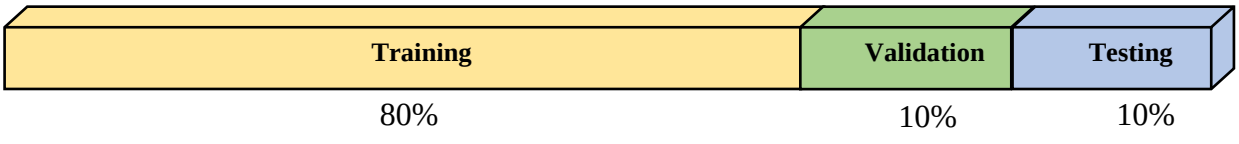


Fig. 3.3 Training, Validation and Testing ratio

To assess the efficacy of the model, three evaluation metrics are employed, namely the mean absolute error (MAE), mean squared error (MSE), and mean fundamental percentage error (MAPE). Metrics are utilized to quantify the disparity between the anticipated values and the actual values. The assessment is conducted on the designated test dataset, and the outcomes are documented.

The implementation presented showcases a time series forecasting model that utilizes a transformer-based approach, integrating both a convolutional neural network and a transformer network. The model's performance on the test set is satisfactory, indicating its potential applicability to diverse time series forecasting tasks. The details about the implementation can serve as a valuable reference for scholars and professionals who intend to utilize this model for their individual time series prediction endeavours.

Comparison of Performance with Baseline Models

The study's performance metrics for a range of models, namely SVR-RBF, SVR-Linear, ARIMA, Boost, LSTM, Transformer, and the proposed CATALYST, are presented in the Table 3.2. The bold text shows the performance of the proposed model. The assessment criteria encompass Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2) score, as you can see in Table 3.2.

The CATALYST model exhibited noteworthy outcomes, as evidenced by its RMSE value of 21.01, MAE value of 13.37, and R2 score of 0.83. The aforementioned metrics demonstrate that CATALYST exhibits higher levels of predictive precision and dependability when compared to alternative models. It is noteworthy that CATALYST exhibits superior performance compared to Boost, LSTM, and Transformer models, which are widely recognized for their efficacy in time series prediction tasks.

Table 3.2 Comparison of different models

MODELS	RMSE	MAE	R2
SVR-RBF[65]	36.46	28.78	0.65
SVR-Linear[66]	42.37	34.24	0.53
ARIMA[67]	55.51	48.28	0.82
BOOST[68]	25.72	12.53	0.84

LSTM[69]	25.71	14.61	0.81
Transformer[70]	24.02	13.92	0.84
CATALYST	**21.01**	**13.37**	**0.89**

3.2.3 Conclusion

The present study introduced a comprehensive methodology for predicting air quality through the utilization of a pre-trained Convolutional Neural Network (CNN) in combination with a Transformer-based approach. The research methodology comprised of three primary phases, namely, data acquisition and pre-processing, learning of short-term contextual features through CNN, and extraction of long-term contextual features using a Transformer.

A dataset of hourly air quality measurements was obtained by utilizing data from five monitoring stations in Delhi, which was subsequently subjected to cleaning and organization. The utilization of the sliding window methodology facilitated the division of the data into input and label blocks, resulting in a more efficient data processing approach and conservation of computational resources. Furthermore, the technique of positional encoding was employed to incorporate sinusoidal positional embeddings into the input data.

During the initial phase of contextual feature acquisition, the input data was presented as an image and processed through a pre-existing convolutional neural network (CNN) to extract pertinent features. This methodology utilized the convolutional neural network's aptitude for detecting spatial patterns and interdependencies present in the dataset. The reduction of dimensionality reduction by converting the data on air quality into a vector image. This enabled the subsequent Transformer model to efficiently carry out temporal modelling efficiently Capture enduring contextual associations; a Transformer network was employed. The Transformer encoder layers were successful in modelling the input data sequence and capturing the requisite context for precise predictions through the integration of self-attention and feedforward neural network operations. The ultimate predictions were generated by the linear layer of the transformer decoder.

The ultimate prediction was attained through the integration of the pre-trained Convolutional Neural Network and the Transformer outputs. The model's performance was enhanced, and overfitting was mitigated through techniques such as dropout and batch normalization. The

evaluation of the model's performance was conducted through the utilization of various metrics, including but not limited to Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2).

The methodology proposed in this study presented a comprehensive framework for predicting air quality. It involved a series of steps including data collection, pre-processing, feature extraction, and prediction analysis. The findings indicated the efficacy of the methodology in precisely predicting atmospheric phenomena. By incorporating both short-term and long-term contextual features, a comprehensive comprehension of the dynamics of air quality can be achieved.

3.3 Significant Outcomes of this Chapter

The significant outcomes of this chapter are as follows:

- To predict PM2.5 concentration levels using remote sensing data through a novel deep learning framework named “CATALYST” (Convolutional and Transformer-based Architecture for Learning Air Quality Spatiotemporal Trends). The proposed model integrates Convolutional Neural Networks (CNNs) for efficient spatial feature extraction and Transformer architectures for capturing long-term temporal dependencies, significantly enhancing predictive accuracy.
- Conducted extensive performance evaluations using metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). The CATALYST model achieved an RMSE of 21.01, MAE of 13.37, and R^2 of 0.89, outperforming state-of-the-art models including ARIMA, LSTM, Boosting algorithms, and standard Transformer approaches.
- Analyzed temporal patterns and seasonal variations in PM2.5 concentrations across Delhi, providing valuable insights into the influence of industrial activities, vehicular emissions, meteorological conditions, and seasonal shifts on air pollution levels. This comprehensive analysis supports data-driven policymaking and targeted pollution control strategies.

The following research studies serve as the foundation for this chapter:

❖ **Abhishek Verma,** Virender Ranga, Dinesh Kumar Vishwakarma, "A novel approach for forecasting PM2.5 pollution in Delhi using CATALYST." Published in Environmental Monitoring and Assessment, Volume 196, Article number 340, (2024).

In this chapter, the BREATH-Net model was introduced, which leverages a hybrid architecture combining a Transformer and a BiLSTM network to accurately forecast NO_2 concentrations in Delhi. By utilizing satellite data and advanced deep learning techniques, the model significantly outperforms traditional forecasting methods, offering valuable insights into the temporal and seasonal patterns of NO_2 pollution.

The next chapter builds on these findings by exploring the model's real-world applications and its potential for integration into broader environmental monitoring systems, focusing on the practical challenges and solutions for deploying such models in urban air quality management.

Chapter 4: BREATH-NET a PRECISION MODEL for NO_2 FORECASTING

4.1 Scope of this Chapter

Air pollution remains a critical environmental and public health concern in the contemporary era of escalating urbanization and industrialization. Among the various air pollutants, nitrogen dioxide (NO_2) significantly contributes to deteriorating air quality, exacerbating respiratory ailments, cardiovascular diseases, and environmental degradation. As one of the most polluted metropolitan regions in the world, Delhi experiences persistent NO_2 pollution due to vehicular emissions, industrial activities, and meteorological influences. Consequently, there is a growing demand for accurate and efficient forecasting models to predict NO_2 concentrations and facilitate effective air quality management strategies.

Conventional forecasting techniques, including statistical regression models and traditional machine learning approaches, often struggle to capture the complex temporal and spatial dependencies inherent in NO_2 pollution data. These methods exhibit limitations in processing the highly dynamic nature of air pollution and integrating multiple influencing factors such as meteorological parameters, emission sources, and seasonal variations. To address these challenges, this study introduces BREATH-Net, a novel hybrid deep-learning framework that integrates a Transformer architecture with a Bidirectional Long Short-Term Memory (BiLSTM) network to enhance the accuracy of NO_2 concentration predictions.

The BREATH-Net model leverages the advantages of Transformer-based attention mechanisms for capturing long-range dependencies and BiLSTM's ability to effectively process sequential data. This combination enhances the model's ability to learn complex temporal patterns and dependencies in NO_2 levels, resulting in more precise and robust forecasts. The study utilizes satellite-based NO_2 data from Sentinel-5P, spanning a period of three years, to train and validate the proposed model. A thorough exploratory data analysis (EDA) is conducted to understand trends and patterns in NO_2 concentrations, followed by pre-processing techniques such as MinMax scaling to optimize the model's performance.

This chapter comprehensively discusses the development, implementation, and evaluation of the BREATH-Net model for NO_2 forecasting in Delhi. The effectiveness of the proposed model is demonstrated through performance comparisons against conventional prediction models such as XGBoost, LSTMs, SVR, and other baseline approaches. The study highlights the superior forecasting capability of BREATH-Net, which achieves a significantly lower Root Mean Square Error (RMSE) of 9.06 and an R^2 score of 0.96, outperforming other state-of-the-art models.

By presenting a robust NO_2 prediction framework, this research aims to contribute to real-time air quality monitoring, support evidence-based policymaking**,** and aid in mitigating the adverse health effects of NO_2 pollution. The insights derived from this study can inform targeted pollution control strategies, optimize emission reduction policies, and foster the development of sustainable urban planning initiatives**.**

4.2 BREATH-Net: A Novel Deep Learning Framework for NO_2 Prediction Using Bi-directional Encoder with Transformer

4.2.1 Abstract

Air pollution poses a significant challenge in numerous urban regions, negatively affecting human well-being. Nitrogen dioxide (NO_2) is a prevalent atmospheric pollutant that can potentially exacerbate respiratory ailments and cardiovascular disorders and contribute to cancer development. The present study introduces a novel approach for monitoring and predicting Delhi's nitrogen dioxide concentrations by leveraging satellite data and ground data from the Sentinel 5P satellite and monitoring stations. The research gathers satellite and monitoring data over 3 years for evaluation. Exploratory data analysis (EDA) methods are employed to comprehensively understand the data and discern any discernible patterns and trends in nitrogen dioxide levels. The data subsequently undergoes pre-processing and scaling utilizing appropriate techniques, such as MinMaxScaler, to optimize the model's performance. The proposed forecasting model uses a hybrid architecture of the Transformer and BiLSTM models called BREATH-Net. BiLSTM models exhibit a strong aptitude for effectively managing sequential data by adeptly capturing dependencies in both the forward and backward directions. Conversely, transformers excel in capturing extensive relationships over extended distances in temporal data. The results of this study will illustrate the proposed model's efficacy in predicting the levels of NO_2 in Delhi. If

effectively executed, this model can significantly enhance strategies for controlling urban air quality. The findings of this research show a significant improvement of RMSE = 9.06 compared to other state-of-the-art models. This study's primary objective is to contribute to mitigating respiratory health issues resulting from air pollution through satellite data and deep learning methodologies.

4.2.2 Proposed Methodology

In this proposed research study, we suggest a methodology to forecast NO_2 (nitrogen dioxide) pollutant levels using a hybrid model that combines a Transformer architecture with a Bi-directional Long Short-Term Memory (BiLSTM) network named as BREATH-Net. The aim is to leverage the strengths of both models to improve the accuracy of NO_2 predictions and contribute to effective air quality management.

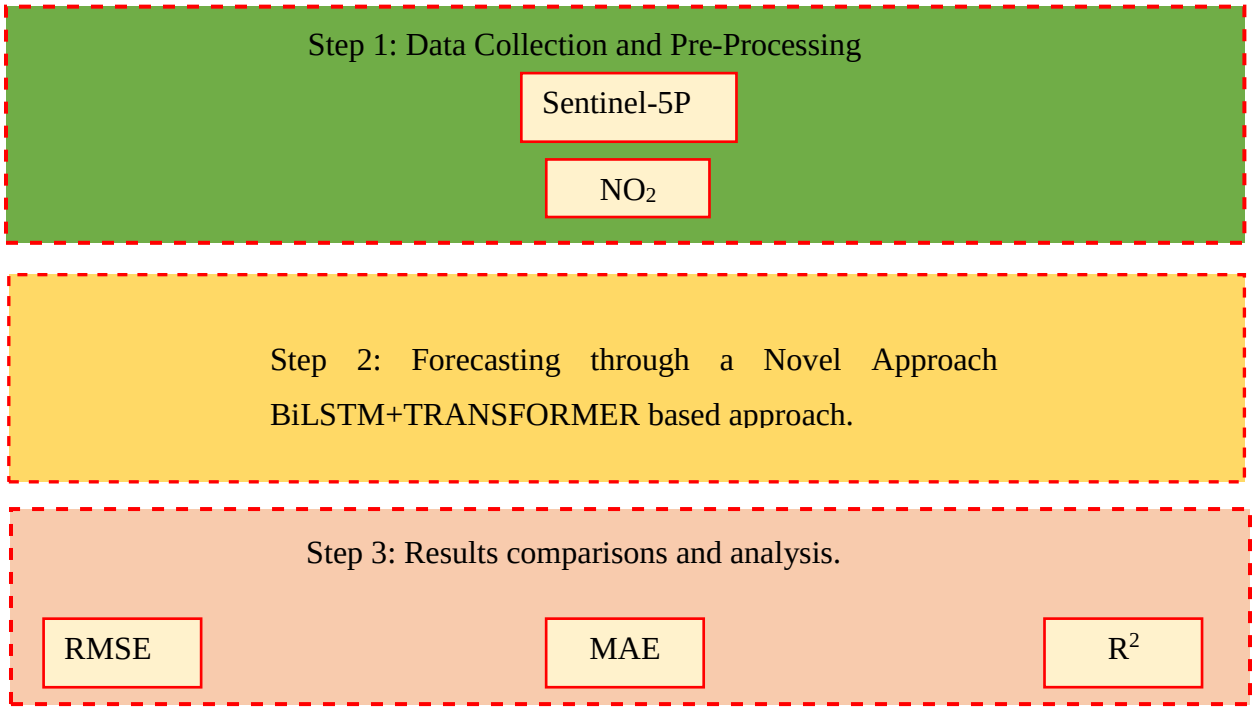


Fig. 4.1 Framework for NO_2 Forecasting

The suggested model architecture effectively combines the attention-based features of the Transformer component with the temporal context modelling capabilities of the BiLSTM network. This hybrid methodology allows the model to proficiently comprehend intricate material connections and generate accurate forecasts about NO_2 pollutant concentrations. The thorough

assessment and analysis of the model's performance are crucial in facilitating efficient air quality management and environmental monitoring.

The proposed approach for forecasting levels of NO_2 pollutants involves the application of a hybrid model BREATH-Net that incorporates a Transformer architecture with a Bi-directional Long Short-Term Memory (BiLSTM) network. This innovative amalgamation's primary objective is to leverage both models' advantages, facilitating a comprehensive analysis of temporal patterns and linkages in the NO_2 data. The input of the model comprises a tensor that represents a sequential series of historical NO_2 concentration values. The tensor exhibits a (num, 1) shape, wherein 'num' denotes the number of preceding time steps taken into account for forecasting. The provided input data facilitates the model's acquisition and comprehension of the temporal patterns and fluctuations in NO_2 pollutant concentrations.

Transformer Component

The Transformer component represents the initial fundamental element of the proposed model. The model utilizes Multi-Head Attention, consisting of four attention heads, each with a critical dimension 32. This mechanism facilitates the model's ability to concurrently focus on various segments of the input sequence, thereby effectively capturing a diverse array of temporal patterns and dependencies inherent in the data. The attention scores in the given equation (4.1) are computed by applying the Softmax function to the dot product of the query matrix (Q) and the key matrix (K). The mathematical formulation presented here serves as a crucial component of the Multi-Head Attention mechanism within the Transformer module of our model. The Softmax function is responsible for normalizing the scores, so ensuring that the model allocates suitable attention weights to various portions of the input sequence in accordance with their significance. Including this phase is of utmost importance to successfully capture a wide range of temporal patterns and relationships within the data.

$$Attention\ Score = SoftMax\left(\frac{QK^T}{\sqrt{d_k}}\right) \tag{4.1}$$

To mitigate the issue of over-fitting, a dropout regularisation technique is implemented after the attention layer, with a dropout rate of 0.1. The utilization of dropout, a method that randomly deactivates a portion of neurons during the training process, improves the model's capacity to

generalize effectively to unfamiliar data. The Dropout operation plays a vital role in the regularisation approach of our model, as seen in the equation (2). During the training process, a certain probability p is used to randomly assign a percentage of the input values to zero. The use of noise in this stochastic process serves the purpose of mitigating overfitting, hence promoting enhanced generalisation of the model to unobserved data.

$$Dropout(x) = \begin{cases} x, & with\ probability\ 1-p \\ 0, & with\ probability\ \ p \end{cases} \tag{4.2}$$

The output of the attention layer is subsequently standardized through Layer Normalisation, employing an epsilon value of 1e-6. Normalization stabilizes the learning process, guaranteeing consistent convergence throughout the training phase. In equation (3), Layer Normalisation (LayerNorm) is a method employed to normalize the activations of individual neurons within a layer in an independent manner. The algorithm computes the arithmetic mean σ and standard deviation μ of the given input values. It then applies scaling and shifting operations to the data in order to achieve a uniform distribution of activations.

$$\text{LayerNorm}(x) = \sigma x - \mu \tag{4.3}$$

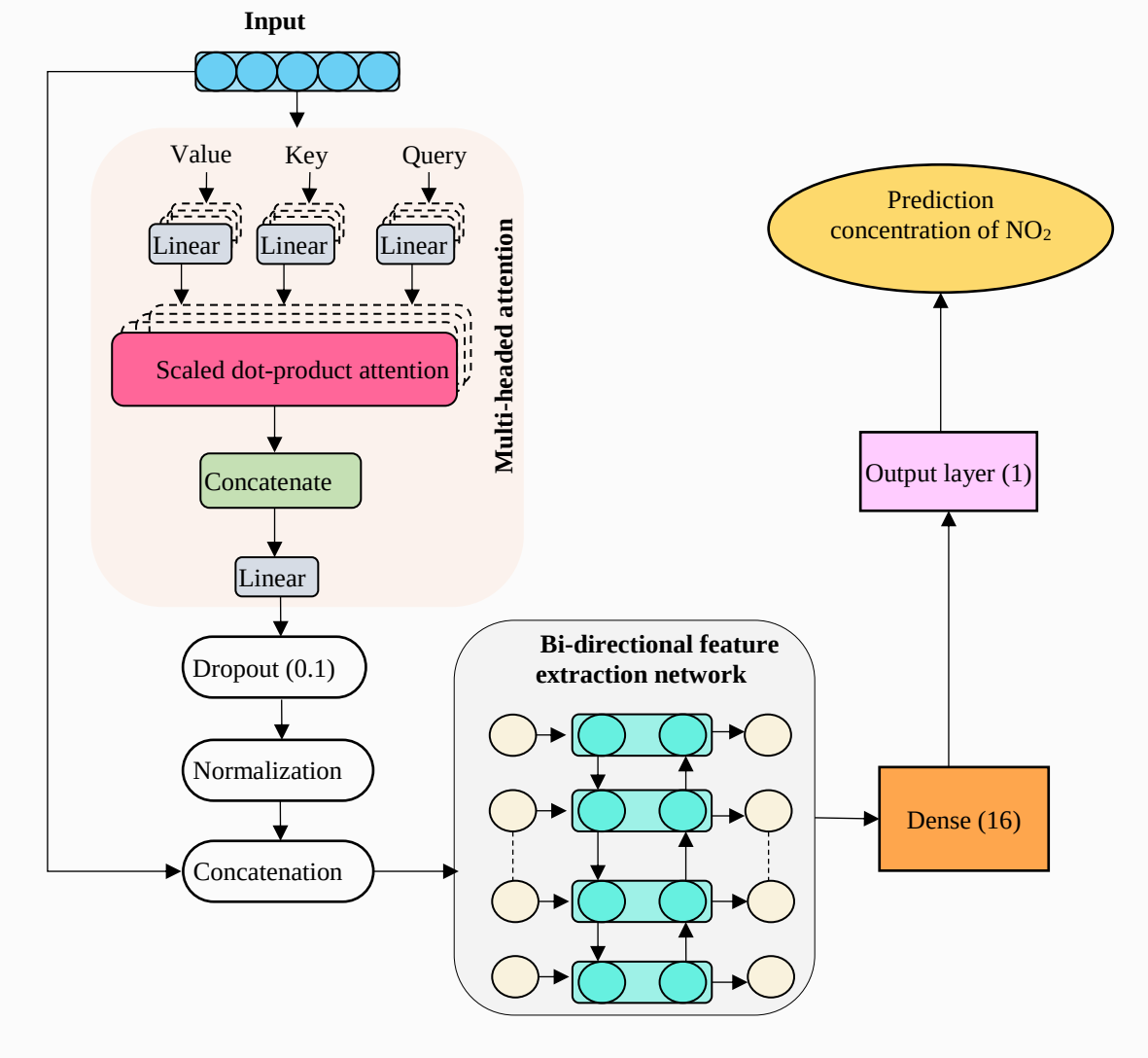


Fig. 4.2 The architecture of Multihead attention and BiLSTM for NO_2 Forecasting

Fused Representation

The fused representation, which incorporates temporal information, is obtained by combining the output of the Transformer model with the initial input sequence. The concatenated sequence is subsequently fed into the BiLSTM layer, enhancing the original time series data by incorporating pertinent contextual information. The given equation (4.4) demonstrates the fusion of the output of the Transformer model X with the original input sequence T at each time step, resulting in the fused representation F. The integration of contextual information acquired by the Transformer with the original time series data results in the production of a novel representation. The use of a fused representation significantly improves the model's capacity to capture temporal patterns and

relationships inherent in the data, thereby rendering it an essential component within the suggested architecture.

$$F = [X, T] \tag{4.4}$$

Dual-directional Long Short-Term Memory (LSTM) Layer

The fused sequence undergoes processing through a Bidirectional Long Short-Term Memory (BiLSTM) layer consisting of 64 units. The bidirectional nature of the BiLSTM enables the model to gather information from both preceding and subsequent time steps, thereby improving its understanding of the temporal patterns in the NO_2 data.

The BiLSTM layer can effectively capture contextual information from preceding and succeeding time steps, allowing the model to discern intricate patterns and temporal dependencies within the data.

Densely Connected Regression Layer

After the BiLSTM processing, the extracted features are further processed using a fully connected Dense layer with 16 units. The dense layer is responsible for transforming the feature representation, characterized by many dimensions, into a format appropriate for regression-based prediction.

The concentration of NO_2 is predicted using a single-neuron output layer, resulting in the ultimate prediction. The output layer of the model generates forecasts for the levels of NO_2 pollutants based on the input time series data and the contextual information extracted by the BREATH-Net components.

Model Training and Hyperparameter Optimization

The hybrid model uses the Adam optimizer, employing a learning rate of 0.001. The Adam optimizer is an adaptive algorithm for optimizing the learning rate, which efficiently updates the parameters of a model during the training process. This results in accelerated convergence and enhanced performance.During the training process, a learning rate annealing scheduler is implemented, whereby the learning rate is systematically reduced by 10 after every 50 epochs. The utilization of the annealing process enhances the convergence of the optimization process by promoting smoother transitions and mitigating the occurrence of overshooting.

Table 4.1 Hybrid Model for Forecasting $\boldsymbol{NO_2}$ Pollutant Levels

Aim: Forecasting $\boldsymbol{NO_2}$pollutant levels is critical for effective air quality management and environmental monitoring

Input: Raw $\boldsymbol{NO_2}$ values, $\boldsymbol{NO_2\_i}$ where, $\boldsymbol{i \leq n}$

Output: Forcasted value for $\boldsymbol{NO_2}$ pollutant level, $\boldsymbol{NO_2\_predicted\_i \in forcasted\ value}$

1. $\boldsymbol{NO_2\_normalized\_i = (NO_2\_i - min(NO_2)) / (max(NO_2) - min(NO_2))}$, normalize $\boldsymbol{NO_2}$ values
2. $\boldsymbol{X\_i = [NO_2\_normalized\_i(t-1), NO_2\_normalized\_i(t-2), \ldots, NO_2\_normalized\_i(t-num)]}$, obtain the input tensor $\boldsymbol{X\_i}$ using different values of normalized $\boldsymbol{NO_2}$ acquired in the previous step

for $\boldsymbol{E} \leftarrow 1$ to *Epochs* do

3. $\boldsymbol{Attention_i = MultiHeadAttention(X_i, num_{heads}, dim_{per_{head}})}$, apply multi-headed attention on the input tensor $\boldsymbol{X_i}$
4. $\boldsymbol{Dropout\_i = ApplyDropout(Attention\_i, dropout\_rate)}$, apply dropout layer on $\boldsymbol{Attention\_i}$ to reduce overfitting
5. $\boldsymbol{LN\_i = LayerNormalization(Dropout\_i)}$, pass the output of the previous layer to the normalization layer
6. $\boldsymbol{Fused\_i = Concatenate(X\_i, LN\_i)}$, fused representation by concatenation of input tensor $\boldsymbol{X_i}$ and the output $\boldsymbol{LN\_i}$ obtained from the layer normalization
7. $\boldsymbol{BiLSTM\_i = BidirectionalLSTM(Fused\_i, num\_units)}$, fused representation passed through the BiLSTM layers
8. $\boldsymbol{Dense\_i = DenseLayer(BiLSTM\_i, num\_units\_dense)}$, output obtained from dense layer
9. $\boldsymbol{NO_2\_predicted\_i = OutputLayer(Dense\_i)}$, forcasted value of NO_2 is obtained by passing through the dense and the output layer

End

4.2.3 Experimental Results and Discussion

This section contains detailed information regarding the dataset utilized during the research, the experimental settings of the proposed framework, and performance assessments.

Dataset Description

The dataset, named "NO_2 Concentration Time Series Data for Delhi City," provides a comprehensive compilation of nitrogen dioxide (NO_2) concentrations observations during a specific period, emphasizing the city of Delhi, as can be seen in Fig. 4.3. The dataset comprises hourly average NO_2 measurements collected at different timestamps ranging from November 25, 2020, to January 24, 2023. Every data point in the collection corresponds to a distinct date and time, together with its respective concentration of NO_2.

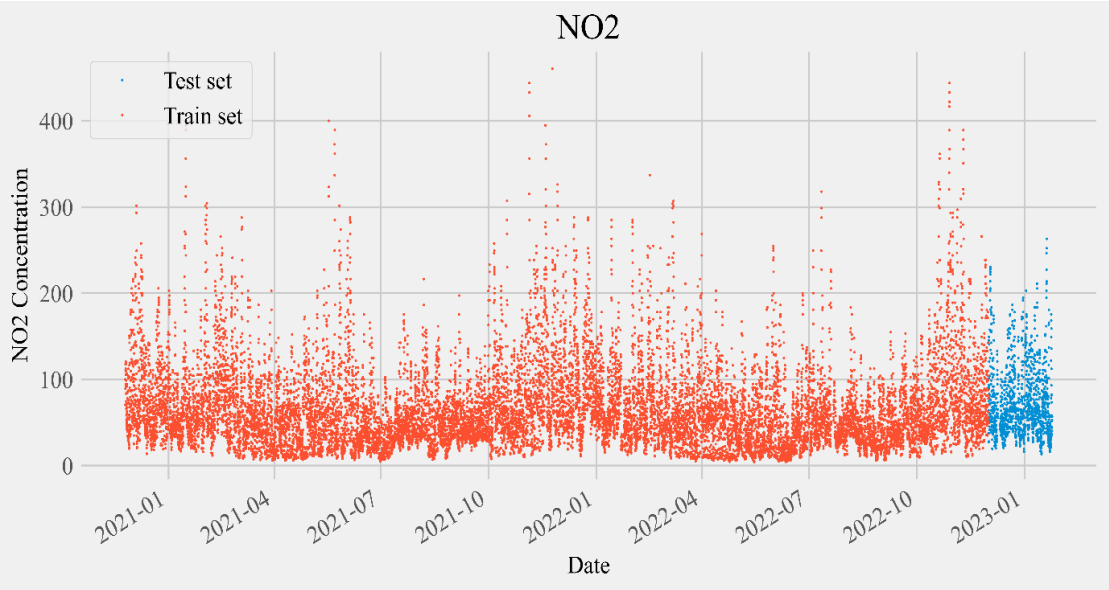


Fig. 4.3 Dataset Timeline

Importance of Nitrogen Dioxide (NO_2)

Nitrogen dioxide (NO_2) is a notable atmospheric contaminant discharged from several origins, encompassing vehicular discharges, industrial operations, and the burning of fossil fuels.

Monitoring nitrogen dioxide (NO_2) levels plays a vital role in evaluating air quality and comprehending air pollution's possible health and environmental consequences.

Temporal Coverage

The temporal coverage of the dataset provides valuable information on the diurnal and seasonal variations in NO_2 concentrations inside Delhi, a region renowned for its elevated pollution levels.

Dataset pre-processing

The 'date' column within the dataset is transformed into a Date Time format to examine temporal patterns comprehensively. Following that, supplementary material characteristics such as 'Year,' 'Month,' 'Day,' and 'Hour' were derived from the 'date' column. Implementing this pre-processing procedure facilitated a more comprehensive comprehension of the temporal patterns and fluctuations in the levels of NO_2 pollution throughout the study. The dataset undergoes a normalization process to guarantee the data's comparability and equity across various scales—the MinMaxScaler function from the sklearn. The pre-processing library is utilized for this objective. The relative relationships between NO_2 values are preserved by scaling the measurements within the range of 0 to 1. The implementation of this normalization procedure plays a vital role in preparing the dataset for subsequent model training. This step is required to make sure that scale-related biases do not interfere with the model's ability to learn from the data. The gathering of a trustworthy and comprehensive dataset for the model's development and analysis later on is made possible by the use of satellite-based remote sensing technologies and the use of data extraction, datetime conversion, and normalization methods. This step is required to make sure that scale-related biases do not interfere with the model's ability to learn from the data. The gathering of a trustworthy and comprehensive dataset for the model's development and analysis later on is made possible by the use of satellite-based remote sensing technologies and the use of data extraction, datetime conversion, and normalization methods.

Implementation Details

The studies are performed on a personal computer (PC) server that is equipped with an NVIDIA QUADRO RTX A5000 graphics card. This graphics card had a memory capacity of 24GB and is equipped with 3042 NVIDIA CUDA cores. The use of machine learning approaches for air quality forecasting necessitates the utilisation of open-source frameworks, such as Pytorch , Sklearn , seaborn and Pandas The open-source Tensorflow library, which can be accessed at

https://github.com/tensorflow/, is utilised for the configuration of deep learning and Transformer models. The given implementation is based on a time series forecasting model that leverages transformers. This model integrates a transformer network with a bidirectional long short-term memory (Bi-LSTM) component and names as BREATH-Net..

Model configuration

The experimental configuration encompassed the integration of a hybrid model that effectively merged a Transformer architecture with a Bi-directional Long Short-Term Memory (BiLSTM) network for the purpose of forecasting nitrogen dioxide (NO_2) pollution levels. The used model is utilised four attention heads inside the Multi-Head Attention mechanism of the Transformer, with each head having a crucial dimensionality of 32. In order to address the issue of overfitting, a dropout regularisation approach is implemented subsequent to the attention layer, with a dropout rate of 0.1. The use of Layer Normalisation was employed to normalise the output of the attention layer. During the training process, the Adam optimizer is utilised with a learning rate of 0.001. Additionally, a learning rate annealing scheduler is applied to progressively decrease the learning rate by a factor of 10 after every 50 epochs. The model underwent training for numerous epochs in order to iteratively update its parameters, hence enhancing its predicting skills.

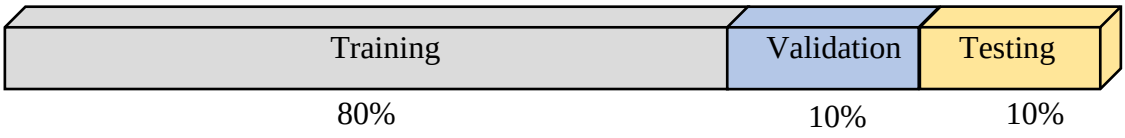


Fig. 4.4 Training, Validation and Testing ratio

Dataset Splitting

The dataset that has undergone pre-processing is divided into three sets, namely the training set , validation set and test set, with a division ratio of 80% ,10% and 10% respectively as seen in Fig. 4.4. The model known as BREATH-Net is compiled using the Adam optimizer, which has a using learning rate annealing technique. The loss function chosen for this model is a mean squared error (MSE). The optimization of model parameters is achieved by iteratively updating them over multiple epochs, enabling the model to acquire knowledge from the available data and enhance its ability to make accurate predictions.

Evaluation Metrics

The quantification of the model's performance was conducted through the use of several assessment measures. The main evaluation criterion employed is the root mean square error (RMSE) equation(4.5), which quantified the precision of the model's forecasts with respect to the observed NO_2 concentration values. Furthermore, the mean absolute error (MAE) in equation(4.6) and the coefficient of determination (R2) in equation (4.7) are utilised as extra metrics in order to offer a more thorough evaluation of the predictive performance of the model. The combination of these measurements provided valuable insights into the model's capacity to provide precise predictions for the amounts of NO_2 pollutants.

$$RMSE = \sqrt{\frac{\sum_{i=0}^{l-1}(z_i - f_i)^2}{l}} \tag{4.5}$$

$$MAE = \frac{\sum_{i=0}^{l-1}|z_i - f_i|}{l} \tag{4.6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{x}(z_i - f_i)^2}{\sum_{i=1}^{x}(z_i - f_i^{mean})^2} \tag{4.7}$$

Within this particular piece, we proceed to disclose the outcomes derived from our comprehensive research investigation. Our primary attention is to evaluate the efficacy of our hybrid model in accurately predicting levels of NO_2 pollutants. Furthermore, we engage in an in-depth analysis and discourse pertaining to the discoveries made during this study. In Table 4.2. ,BREATH-Net is highlighted in bold and demonstrates superior performance over all other relevant models, with the predicted values presented in a detailed illustration for comparison.

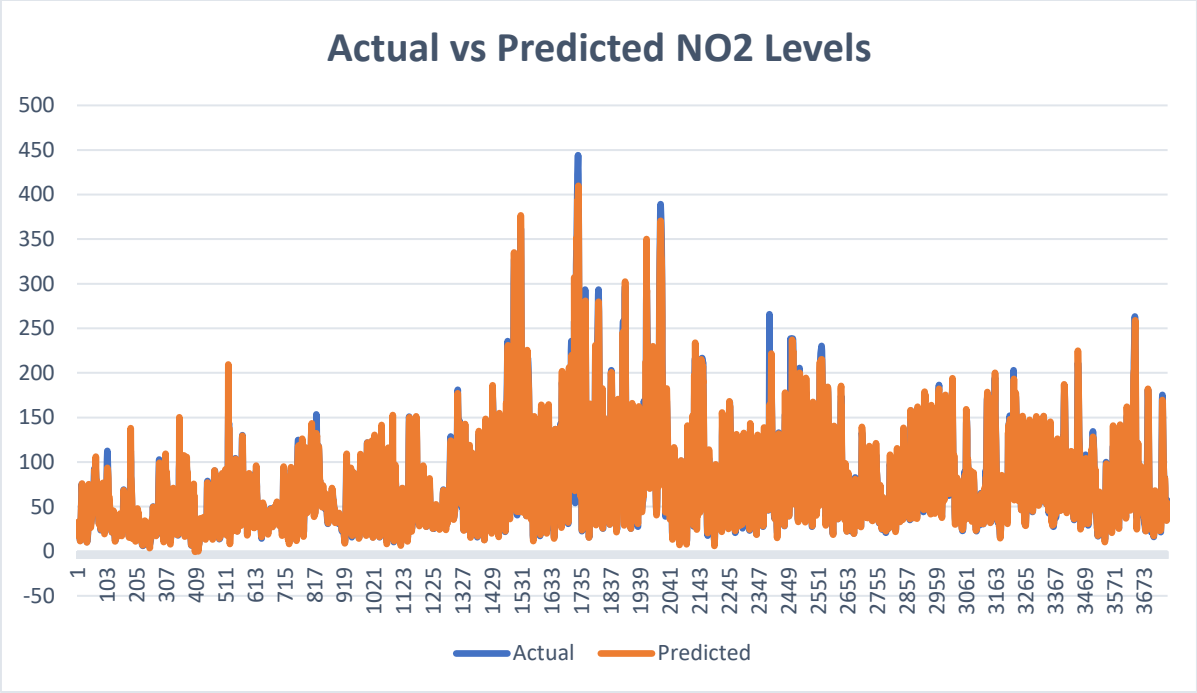


Fig. 4.5 Predicted values vs Actual values

Table 4.2 Comparison of Performance with Baseline Models

MODELS	RMSE	MAE	R2
XGB Boost[68]	12.01	6.62	0.95
SVR-Linear[71]	19.45	15.58	0.86
GRU +LSTM[69]	12.27	7.34	0.94
FB Prophet[72]	63.07	44.48	0.11
LSTM[73]	13.78	7.45	0.91
Transformer[74]	20.02	10.92	0.89
BREATH -Net	**9.06**	**5.11**	**0.96**

Among the different models assessed for NO_2 forecasting, "BREATH-Net" demonstrates superior performance, establishing itself as the most prominent contender. The predictive accuracy and fit of "BREATH-Net" are noteworthy, as evidenced by its exceptionally low root mean square error (RMSE) of 9.06, the lowest mean absolute error (MAE) of 5.11, and an impressive coefficient of determination (R2) score of 0.96. These results indicate that "BREATH-Net" performs exceptionally well in accurately predicting the outcomes and effectively capturing the patterns in the data. By utilising sophisticated neural network architecture and employing innovative

techniques, this model demonstrates exceptional proficiency in capturing the complex patterns and dynamics exhibited by NO_2 concentration data. The exceptional performance of this model establishes it as the foremost option for NO_2 prediction, providing researchers and practitioners with a dependable instrument for making well-informed decisions in the realm of air quality control and management, surpassing other models in a comprehensive manner.

Predicted Values Analysis and EDA

In order to comprehend the temporal patterns and seasonal fluctuations of this important air pollutant, we thoroughly examined Nitrogen Dioxide (NO_2) concentrations in the Delhi region for this study. Box plots, a potent graphic technique that effectively depicts the distribution and statistical measurements of NO_2 concentrations throughout several dates and seasons, were used to illustrate our findings graphically. Each box plot in our visualization offers a different angle on the data, analysing each weekday concerning several seasons. The interquartile range (IQR), which encompasses the middle 50% of the data, is shown by the centre box of the graphic. The horizontal line inside the box shows the median concentration of NO_2, while the bottom and top margins of the box represent the 25th and 75th percentiles, respectively. The whiskers that protrude from the boxes show the dispersion and fluctuation of NO_2 concentrations. The top whisker shows the most outstanding value within 1.5 times the IQR, while the minimum value within 1.5 times the lower whisker shows the IQR. Outliers are extreme data points independently shown as distinct points outside the whiskers to shed light on unusual occurrences or noteworthy abnormalities. The box plot analysis provides fascinating new information on NO_2 pollution's spatiotemporal trends. For instance, throughout the Autumn and Spring seasons, we saw greater NO_2 concentrations on Mondays, with mean values of around 77.21 g/m3 and 80.65 g/m3, respectively as seen in Fig. 4.6.

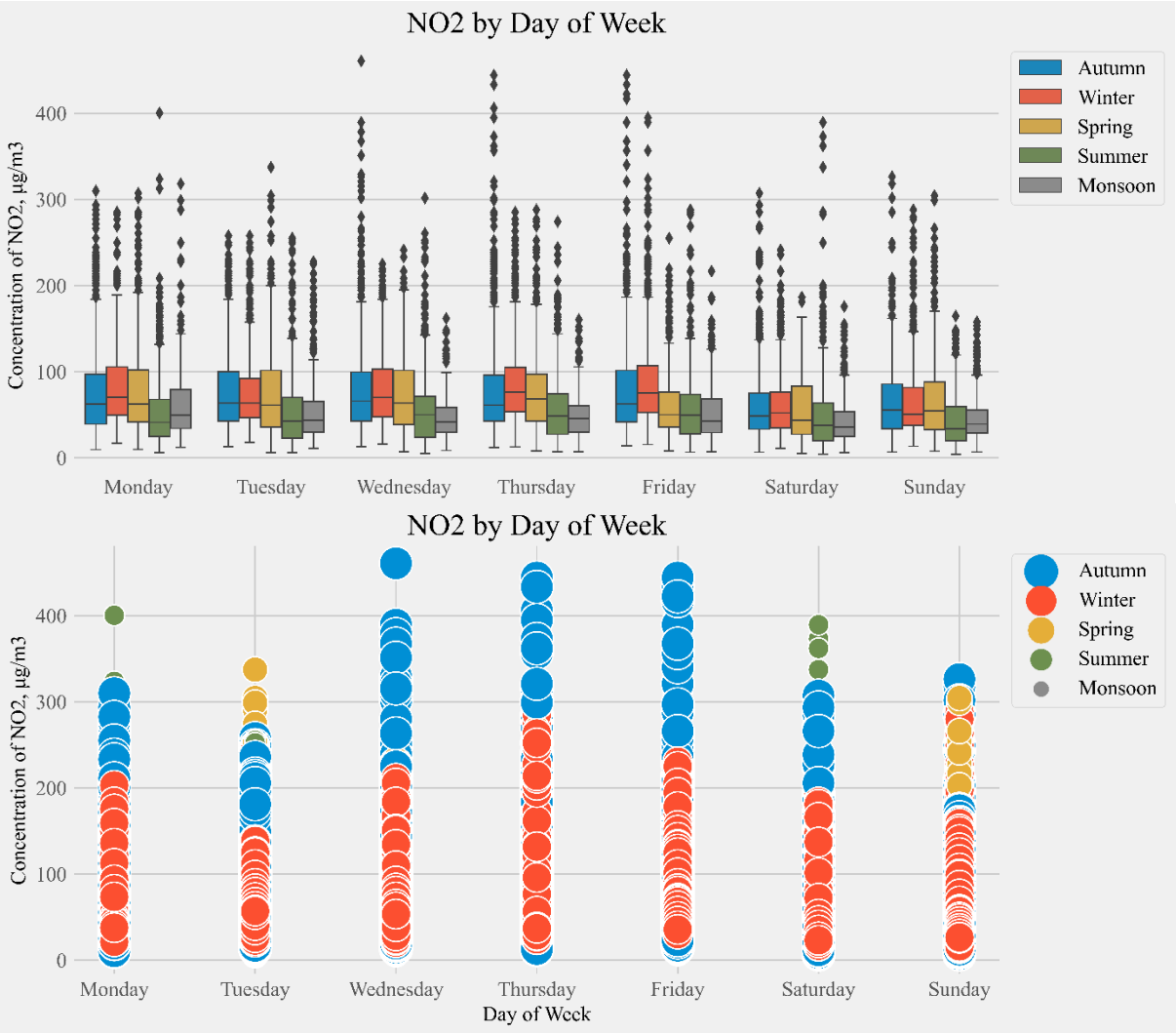


Fig. 4.6 Box and whisker and Dot plot showing NO_2 Conc. day-wise

On the other hand, during the Summer, the mean NO_2 concentration on Mondays was around 52.91 g/m3, which was lower in comparison. These data point to significant weekday-based changes in NO_2 concentrations, which may affect our knowledge of how industrial processes, traffic patterns, and weather patterns affect air quality. The seasonal differences also highlight the importance of the atmosphere and the sources of emissions at various periods of the year. Policymakers, academics, and stakeholders may get essential insights into the temporal dynamics of NO_2 pollution in the Delhi region by using the box plot visualization which is clearly visible in Fig. 4.8.

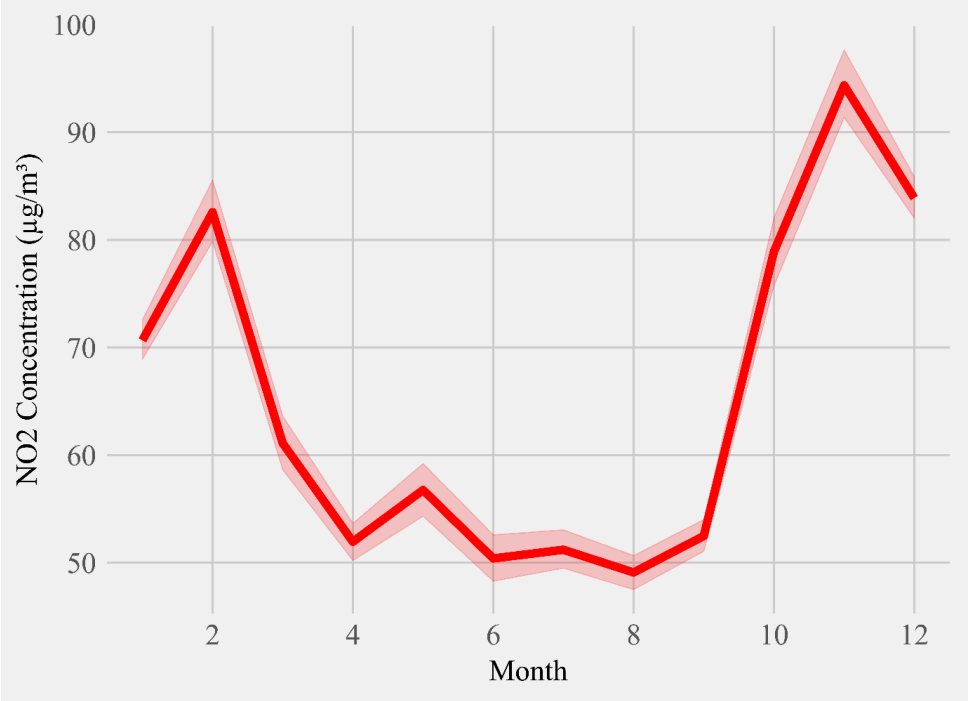


Fig. 4.7 Monthly trend NO_2 conc. level

In order to comprehend the temporal patterns and seasonal fluctuations of this important air pollutant, we thoroughly examined Nitrogen Dioxide (NO_2) concentrations in the Delhi region for this study. Box plots, a potent graphic technique that effectively depicts the distribution and statistical measurements of NO_2 concentrations throughout several dates and seasons, were used to illustrate our findings graphically.

This study comprehensively investigates the levels of Nitrogen Dioxide (NO_2) in the Delhi area for a period of 12 months. The graphical representations employed in this study serve as visual aids to depict the patterns and fluctuations in nitrogen dioxide (NO_2) levels, utilising data derived from monthly observations. The main objective of our study was to provide valuable insights for air quality control initiatives by examining the temporal trends of NO_2 pollution and identifying potential periods of heightened risk.

The line figure illustrates the concentration trend of NO_2 over the course of many months, emphasising notable seasonal variations. Significantly, the concentrations of NO_2 exhibited a noticeable increase throughout the month of November, characterised by a mean value of 94.36 g/m3 and a reported maximum concentration of 460.62 g/m3. In contrast, there was a notable

decline in NO_2 concentrations during the months of April and May, with average values of 51.94 g/m3 and 56.75 g/m3, respectively.

In addition, we noticed that NO_2 levels varied with the seasons, with winter (December to February) showing greater NO_2 concentrations than other times of the year. Due to increasing emissions from heating and vehicular activity, the mean NO_2 concentration for this period was around 79.40 g/m3. In contrast, NO_2 levels were much lower during the monsoon season (July to September), with a mean value of 50.99 g/m3, possibly due to rain-dispersing pollutants and decreased industrial emissions.

The study also focuses on the graph effectively depicts the fluctuations, showcasing a prominent initial concentration of 42.78 $\mu g/m^3$ at the onset (0-hour). Following this, the concentration exhibited a gradual and consistent rise, culminating at the 6th hour with an impressive peak value of 69.65 $\mu g/m^3$ as it can be observed in Fig. 4.8 . Subsequent to this apex, a noticeable decrease occurred, resulting in a significant drop to a minimum level of 18.92 $\mu g/m^3$ by the eighth hour. Remarkably, there was a subsequent reversal in the observed pattern, characterised by a gradual increase towards the duration of 14.5 hours, accompanied by a surge in PM2.5 concentrations to approximately 114.98 $\mu g/m^3$. As the passage of time continued, specifically as it neared the 20th hour, a significant decline was observed, eventually reaching a steady state at a concentration of 60 $\mu g/m^3$. Ultimately, after the completion of the 24-hour duration, the concentration reverted back to its initial value of 42 $\mu g/m^3$. The comprehensive examination conducted on an hourly basis highlights the ever-changing characteristics of air quality, which is marked by noticeable high and low points. As a result, this analysis provides significant and indispensable knowledge for the purposes of environmental monitoring and the development of management strategies.

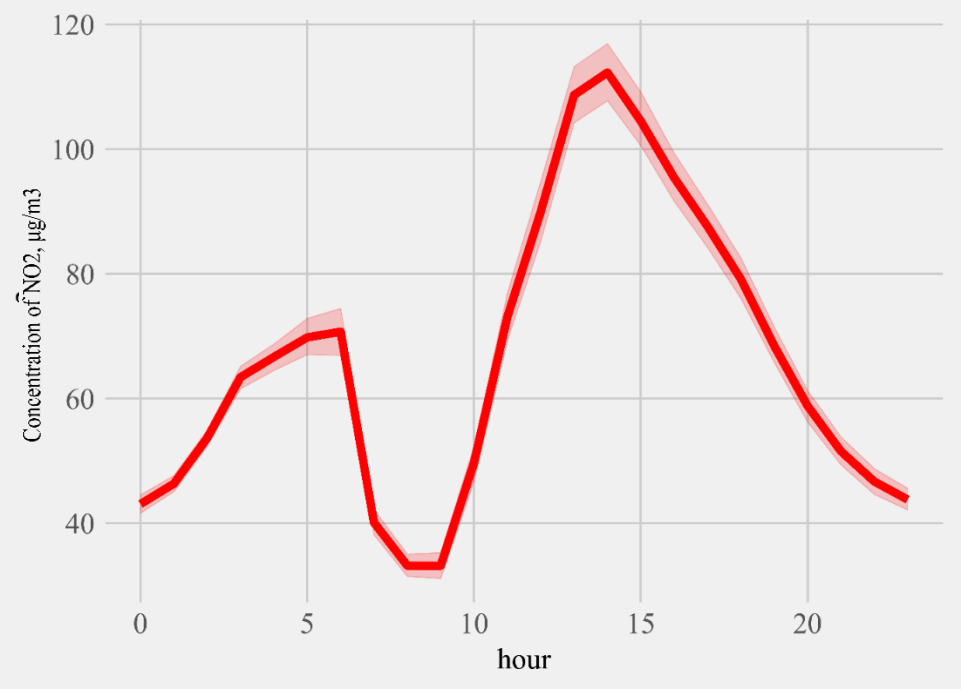


Fig. 4.8 Hourly depiction of NO_2 levels

Policymakers, researchers, and environmental authorities may thoroughly study the NO_2 concentration trend to develop practical solutions for reducing air pollution in the Delhi region. We may apply targeted interventions and restrictions to minimize NO_2 emissions at peak times and support sustainable air quality management by comprehending the temporal patterns and seasonal changes of NO_2.

To fully understand air pollution dynamics, this study emphasizes the importance of ongoing monitoring and analysis of air quality data. These results are essential in creating evidence-based policies and strategies to safeguard public health and improve the general well-being of the population in the Delhi region as we work to create a cleaner and healthier environment.

To comprehend the temporal patterns and seasonal fluctuations of this vital air pollutant, we thoroughly examined nitrogen dioxide (NO_2) concentrations in the Delhi region for this study. Box plots, a potent graphic technique that effectively depicts the distribution and statistical measurements of NO_2 concentrations throughout several dates and seasons, were used to illustrate our findings graphically.

In this study, a comprehensive analysis was conducted on the levels of Nitrogen Dioxide (NO_2) in the Delhi region over a span of 12 months . The study employed monthly measurements and employed graphical representations to depict the patterns and fluctuations in NO_2 levels visually. The main objective of our study was to provide valuable insights for air quality control initiatives by examining the temporal variations of NO_2 pollution and determining potential periods of heightened risk.

The line plot visually represents the concentration trend of NO_2 over the course of many months, emphasizing notable seasonal patterns. Notably, NO_2 concentrations increased noticeably in November, with a mean value of 94.36 g/m3 and a maximum-recorded concentration of 460.62 g/m3. On the other hand, the NO_2 concentrations significantly decreased in April and May, with mean values of 51.94 g/m3 and 56.75 g/m3, respectively same can be seen in Fig. 4.9.

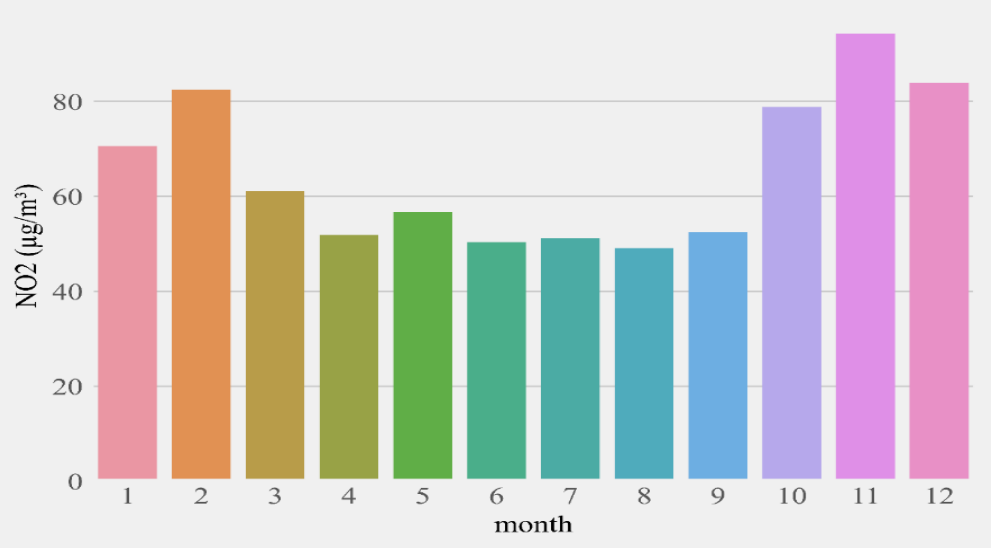


Fig. 4.9 Depiction of NO_2 levels based on months

In addition, we noticed that NO_2 levels varied with the seasons, with winter (December to February) showing greater NO_2 concentrations than other times of the year. Due to increasing emissions from heating and vehicular activity, the mean NO_2 concentration for this period was around 79.40 g/m3. In contrast, NO_2 levels were much lower during the monsoon season (July to September), with a mean value of 50.99 g/m3, possibly due to rain-dispersing pollutants and decreased industrial emissions as it is depicted in Fig. 4.9.

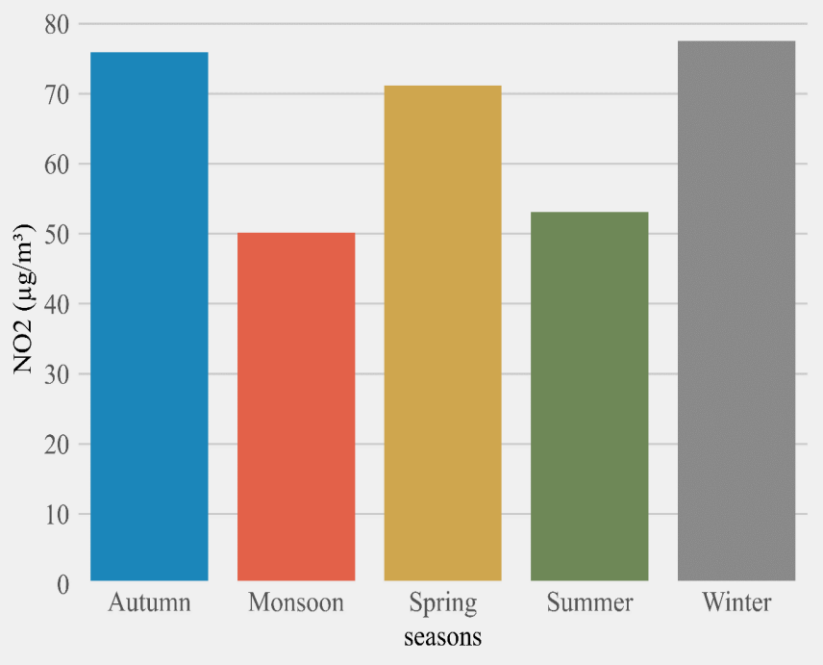


Fig. 4.10 Season wise level of NO_2 levels

Policymakers, researchers, and environmental authorities may thoroughly study the NO_2 concentration trend to develop practical solutions for reducing air pollution in the Delhi region. We may apply targeted interventions and restrictions to minimize NO_2 emissions at peak times and support sustainable air quality management by comprehending the temporal patterns and seasonal changes of NO_2.

To fully understand air pollution dynamics, this study emphasizes the importance of ongoing monitoring and analysis of air quality data. These results are essential in creating evidence-based policies and strategies to safeguard public health and improve the general well-being of the population in the Delhi region as we work to create a cleaner and healthier environment.

4.2.4 Conclusion

This study introduced a novel hybrid model that integrated a Transformer architecture with a Bidirectional Long Short-Term Memory (BiLSTM) network to forecast Delhi's nitrogen dioxide (NO_2) pollutant concentrations. The proposed model effectively utilized the advantages of both architectures, allowing it to capture intricate temporal patterns and dependencies within the NO_2 data. The study's presentation of experimental outcomes offered evidence supporting the effectiveness of the suggested technique in generating accurate forecasts and demonstrating

promising potential for air quality management. The investigation examined the temporal patterns and seasonal variations of nitrogen dioxide (NO_2) concentrations through graphical representations, including box and line plots. The results indicated notable fluctuations in NO_2 concentrations based on weekdays and seasons, providing insights into the impact of industrial activities, traffic flow, and meteorological factors on atmospheric conditions. This information has the potential to provide valuable guidance to policymakers, researchers, and stakeholders in the formulation of focused interventions and policies aimed at mitigating NO_2 emissions during periods of heightened risk, thereby enhancing the management of air quality. The performance of the hybrid model was assessed by employing the root mean square error (RMSE) metric, which measured the accuracy of the model's predictions relative to the true values. The findings suggested that the proposed methodology exhibits promising results in forecasting levels of NO_2 pollutants. The thorough evaluation and analysis of the model's performance are crucial in facilitating efficient air quality management and environmental monitoring.

4.3 Significant Outcomes of this Chapter

The significant outcomes of this chapter are as follows:

- To predict nitrogen dioxide (NO_2) concentration levels using satellite data through a novel deep learning framework named BREATH-Net (Bi-directional Encoder with Transformer for NO_2 Prediction). The proposed model integrates two primary components: Bi-directional Long Short-Term Memory (BiLSTM) networks for capturing sequential dependencies and Transformer architecture for leveraging attention mechanisms to model long-range temporal relationships, significantly enhancing predictive accuracy.
- Implemented robust data preprocessing techniques such as MinMaxScaler for normalization, and optimized model training using the Adam optimizer with learning rate annealing, improving the model's convergence and generalization on unseen data.
- Conducted extensive performance evaluations, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). BREATH-Net achieved an RMSE of 9.06, MAE of 5.11, and R^2 of 0.96, outperforming other state-of-the-art models like XGBoost, SVR-Linear, GRU+LSTM, and standalone Transformer architectures.

- Performed comprehensive Exploratory Data Analysis (EDA) to examine temporal patterns and seasonal variations in NO_2 concentrations across Delhi, revealing significant weekday and seasonal fluctuations, thus providing valuable insights into the impact of industrial activities, vehicular emissions, and meteorological factors on air quality.
- Assessed the robustness and applicability of the BREATH-Net model in real-world urban air quality management, demonstrating its potential to support policy-making for pollution control and mitigation of respiratory health risks associated with NO_2 exposure.

The following research studies serve as the foundation for this chapter:

❖ Abhishek Verma, Virender Ranga, Dinesh Kumar Vishwakarma, "BREATH-Net: a novel deep learning framework for NO_2 prediction using bi-directional encoder with transformer." Published in Environmental Monitoring and Assessment, Volume 196, Article number 340, (2024), IF – 3.0.

This chapter presents the BREATH-Net model, a hybrid deep learning framework combining Transformer and BiLSTM architectures for accurate NO_2 forecasting, with performance evaluations compared to existing models.

The next chapter introduces Arctic-Net, a hybrid model designed for efficient sea ice classification using SAR images, focusing on its innovative architecture, training techniques, and exceptional performance in environmental monitoring.

Chapter 5: ARCTIC-NET: HYBRID DEEP LEARNING MODEL FOR AUTPOMATED SEA-ICE CLASSIFICATION

5.1 Scope of this Chapter

In the era of rapid climate change and increasing concerns about Arctic and Antarctic ice dynamics, accurate sea ice classification has become a critical component of environmental monitoring, climate modelling, and maritime navigation. The ability to efficiently classify and monitor sea ice using Synthetic Aperture Radar (SAR) images is essential for understanding global climate patterns, optimizing shipping routes, and mitigating risks associated with Arctic exploration. Traditional classification methods, such as statistical modelling and classical machine learning techniques, often struggle to effectively analyse the complex spatial and temporal variations in SAR data, limiting their accuracy and generalization capabilities.

To address these challenges, this chapter presents Arctic-Net, a novel hybrid deep learning framework designed to enhance the efficiency and precision of sea ice classification. Arctic-Net integrates Convolutional Neural Networks (CNNs) with attention-based mechanisms, leveraging the strengths of both approaches to capture local texture features and global contextual dependencies within SAR images. This hybrid methodology allows the model to outperform existing state-of-the-art classifiers, including DenseNet, ResNext, and Swin Transformer, by achieving superior accuracy, precision, and computational efficiency.

This chapter provides a comprehensive discussion of the Arctic-Net framework, including its architectural components: Adaptive Convolutional Encoder (ACE), Spatial Transposer Encoder (STE), and Hierarchical Transpose Attention (HTA). It explores the dataset utilized for training and validation, detailing preprocessing techniques such as image normalization, augmentation, and stratified sampling to improve model robustness. The experimental setup, including hardware configurations and hyperparameter optimization, is also elaborated to ensure reproducibility.

The performance evaluation of Arctic-Net is conducted through rigorous comparisons against benchmark models, demonstrating its ability to achieve a classification accuracy of 0.93, a

precision of 0.91, and an F1-score of 0.91. Additionally, the chapter highlights qualitative analyses using visualization techniques such as Layer CAM, which provides interpretability by illustrating how the model distinguishes between different ice categories.

By introducing Arctic-Net, this research contributes to advancing automated sea ice classification, enabling more efficient and scalable solutions for remote sensing applications. The insights derived from this study can aid in the development of real-time sea ice monitoring systems, support climate policy decisions, and facilitate safer maritime operations in polar regions. Future advancements could explore the integration of multimodal remote sensing data, self-supervised learning techniques, and real-time deployment strategies to further improve the applicability of Arctic-Net in operational environments.

5.2 Arctic-Net: A Hybrid Convolutional and Attention-Based Model for Efficient Sea Ice Classification Using SAR Images

5.2.1 Abstract

Sea ice classification accuracy is crucial for climate research, marine navigation, and environmental monitoring. This study presents Arctic-Net, a unique hybrid model that improves sea ice categorization using SAR pictures by combining convolutional neural networks (CNNs) with attention processes. The Adaptive Convolutional Encoder (ACE), the Spatial Transposer Encoder (STE), and the Hierarchical Transpose Attention (HTA) mechanism make up the three main parts of the model. Together, these elements extract global and local information with high efficiency, allowing for accurate sea ice classification with low computing costs. On a dataset of 4,000 SAR pictures, the Arctic-Net model achieves an accuracy of 0.93, precision of 0.91, and F1-score of 0.91, outperforming many state-of-the-art models, such as DenseNet, ResNext, and Swin Transformer. This makes operational sea-ice monitoring and classification tasks in resource-constrained contexts a strong solution. The paper's conclusion includes a review of potential future research avenues and industrial uses for Arctic-Net in real-time sea ice monitoring systems.

5.2.2 Proposed Methodology

The proposed sea-ice classification method is a novel hybrid framework that employs the advantages of Convolutional Neural Networks (CNNs) and attention processes to improve the

extraction of features and classification accuracy. The model is precisely engineered to effectively handle and examine SAR images, collecting spatial texture features and backscattering information. These factors are crucial for precisely distinguishing various forms of sea ice. Table **5.1** presents the training process for the Arctic-Net model

Table 5.1 Pseudocode of the Proposed "Arctic -Net" Model

Input: Dataset= $\{X_i, Y_i\}_{i=1}^{N}$, $X_i \in R^{3\times224\times224}$ illustrating input images & $Y_i \in \{0,1,2,3,4,5,6\}$ as corresponding labels
Model parameters θ
Batch Size β
Epoch $\mathcal{E}$
Learning Rate Lr
After n epochs, Learning Rate Decay Factor γ, where $\gamma \in [0, 1]$
Output: Trained Arctic-NET model for Sea Ice assessment

Initialize θ and the adaptive weights α
for $i = 1$ ... ... $\mathcal{E}$ **do** (Training for $\mathcal{E}$ epochs)
  for $j = x_1$ ... ... x_β **do** (Iterate through each batch β within the epoch)
    $(x_\beta, y_\beta)\rho$ (Randomly select one batch with a size of β)
    $\hat{y}_b = (x_\beta; \theta)$ (Compute posterior probability for each input sequence)
    L_{CE} = $cross\ entropy\ loss\ (y, y_\beta)$ (Calculate cross-entropy loss)
    $\theta \Leftarrow \theta - Lr\Delta_\theta L_{CE}(y_F, y)$ (Optimize model parameters by minimizing. computing loss{ θ } cross-entropy loss L_{CE} using backpropagation)
    if ($i\%n == 0$)
      end for
If i % n=0 then
  $Lr \leftarrow Lr \times \gamma$ (Decay learning rate after every 'n' epochs)
  end for
return None

Arctic-Net Framework

The Arctic-Net model consists of three primary components: the Adaptive Convolutional Encoder (ACE), the Spatial Transposer Encoder (STE), and the Hierarchical Transpose Attention (HTA) mechanism. The purpose of these components is to enhance the process of extracting features and encoding global context in a computationally efficient manner. The presence of this structure is essential for developing a solid deep neural network when resources are constrained.

The ACE module utilizes depth-wise separable convolutions with dynamic kernel widths ranging from 3 to 9 to minimize computational effort while preserving feature efficacy. This concept improves the representation of local features using flexible kernels, resulting in better

performance than fixed alternatives. ACE incorporates Layer Normalization (LN) and Gaussian Error Linear Unit (GELU) activations to provide accurate and effective non-linear feature mapping. In addition, ACE has a skip connection to guarantee seamless information transmission inside the network.

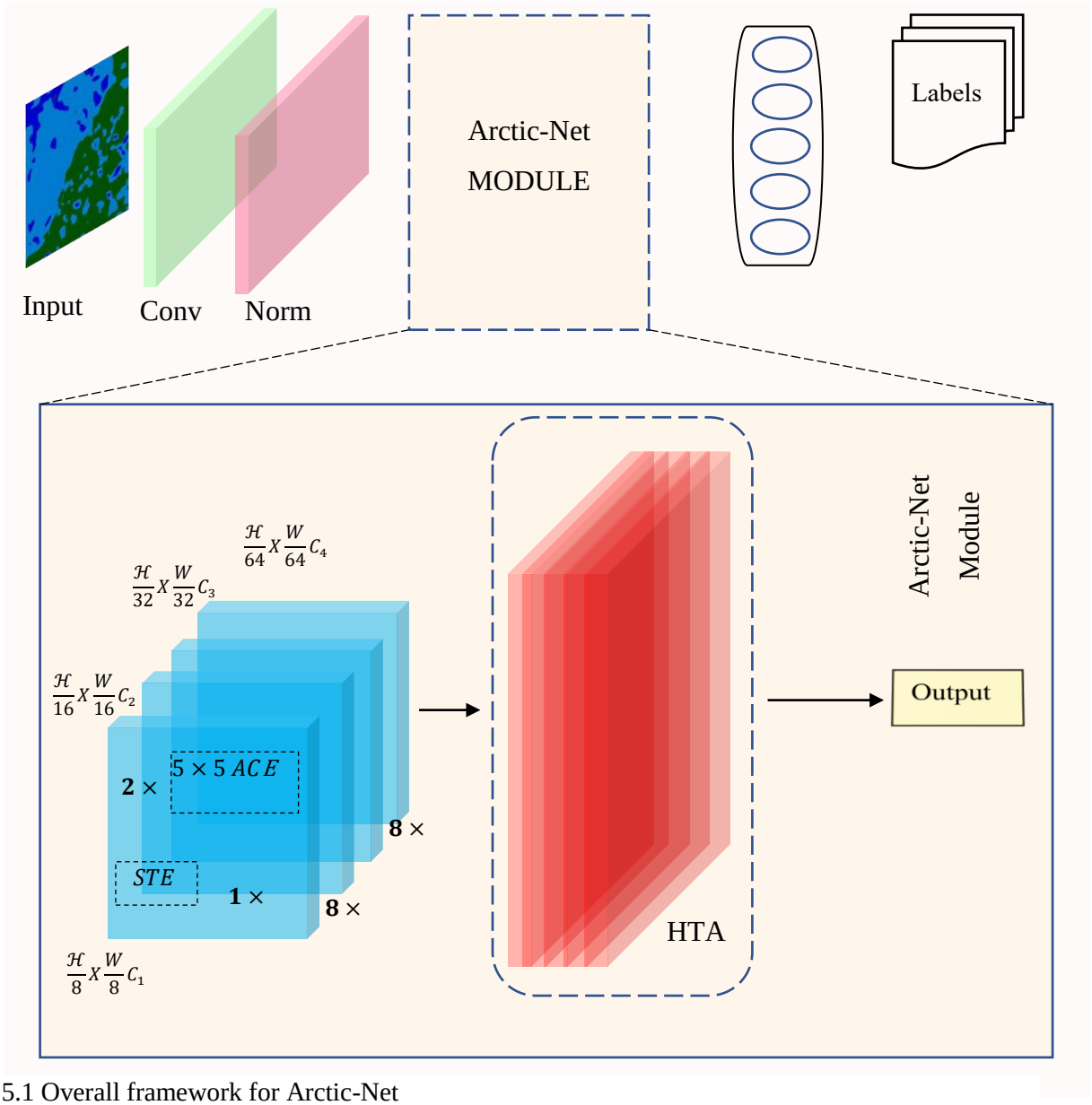


Fig. 5.1 Overall framework for Arctic-Net

The STE module is designed to teach the acquisition of adaptable multiscale feature representations and improve comprehension of global context. The input tensor is divided into subsets, and each subgroup is processed using depth-wise convolutions. The results are then

combined to capture both detailed and global representations. This approach guarantees the creation of an efficient network in terms of its parameters, resulting in a reduced computing load compared to conventional techniques.

The HTA method in Arctic-Net is a fundamental invention integrating components from Vision

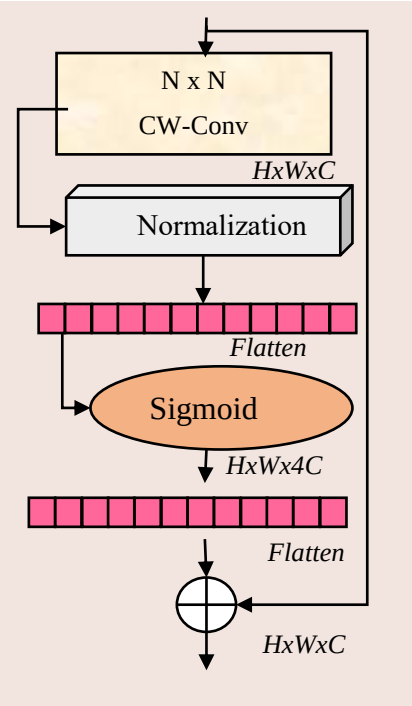


Fig. 5.2 Adaptive Conv. Encoder

Transformers (ViTs) and Convolutional Neural Networks (CNNs). The HTA architecture begins by evaluating an input image with dimensions of 224×224. It next divides the image into smaller patches, either 14×14 or 7×7 in size.

The patches are transformed into linear embedding tokens and then inputted into a series of HTA blocks. These blocks are specifically intended to encode both spatial and channel information. The HTA block is composed of two main modules: Permute-MLP and Channel-MLP. The Permute-MLP module processes spatial information, whereas the Channel-MLP module encodes channel information. The Weighted Permute-MLP improves this process by adaptively modifying the significance of different branches through divided attention, enhancing network performance.

Integrating these components into the Arctic-Net model architecture simplifies the creation of a

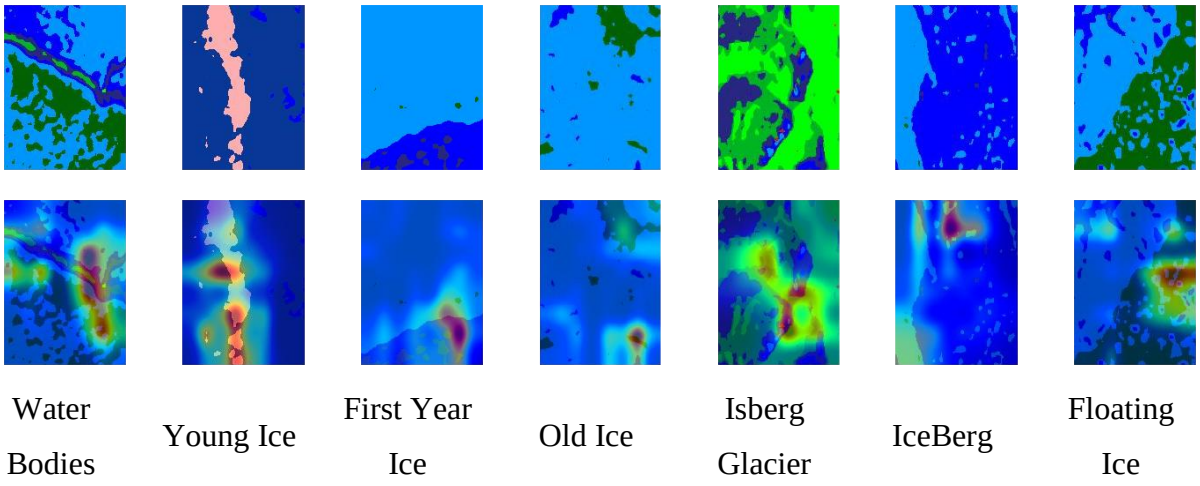


Fig. 5.3 Layer CAM visualization of a SAR image from the Sentinel-1 product (Wishart), with the corresponding layer output.

precise and economical deep neural network. Combining the ACE, STE, and HTA modules allows for complete feature extraction and global context representation while minimizing computing complexity. By employing this comprehensive method, Arctic-Net guarantees exceptional precision and effectiveness, rendering it an ideal choice for vision tasks on devices with limited resources.

The layer CAM representation in graphically emphasizes the model's attention on particular input portions that significantly affect the categorization decision. Additionally, it exposes the distinct characteristics present in these areas, such as textures, forms, or edges, which the model depends on to formulate its forecast. Arctic-Net offers a robust and expandable solution for complex deep learning tasks by utilizing innovative feature extraction methods and attention mechanisms.

Adaptive Convolutional Encoder (ACE)

The adaptable Convolutional Encoder (ACE) presents a new method for representing features using depth-wise separable convolutions with dynamic and adaptable kernel sizes; it can also be observed in Fig. 5.2 .ACE is a two-layer structure that improves the representation of local features by using adaptable N×N kernels. These kernels have sizes of 3, 5, 7, and 9 at different phases. This design decision enhances performance compared to static kernel alternatives and incorporates

conventional Layer Normalisation (LN) and Gaussian Error Linear Unit (GELU) activations for reliable non-linear feature mapping.

Unlike typical Convblocks, ACE incorporates a skip connection to guarantee smooth information transmission within network topologies. This novel paradigm has exhibited exceptional performance, . The following equations precisely specify the ACE.

$$z_{i+1} = z_i + Flatten_G(Flatten(N(CW(z_i))) \quad (5.1)$$

The equation (1) demonstrates the relationship between the input feature maps z_i denotes the input feature maps of shape H×W×C, $Flatten_G$ is a point-wise convolution layer followed by GELU, CW is $k \times k$ Channel-wise convolution, N is a normalization layer, and z_{i+1} denotes the output feature maps of the ACE.

Spatial Transposer Encoder (STE)

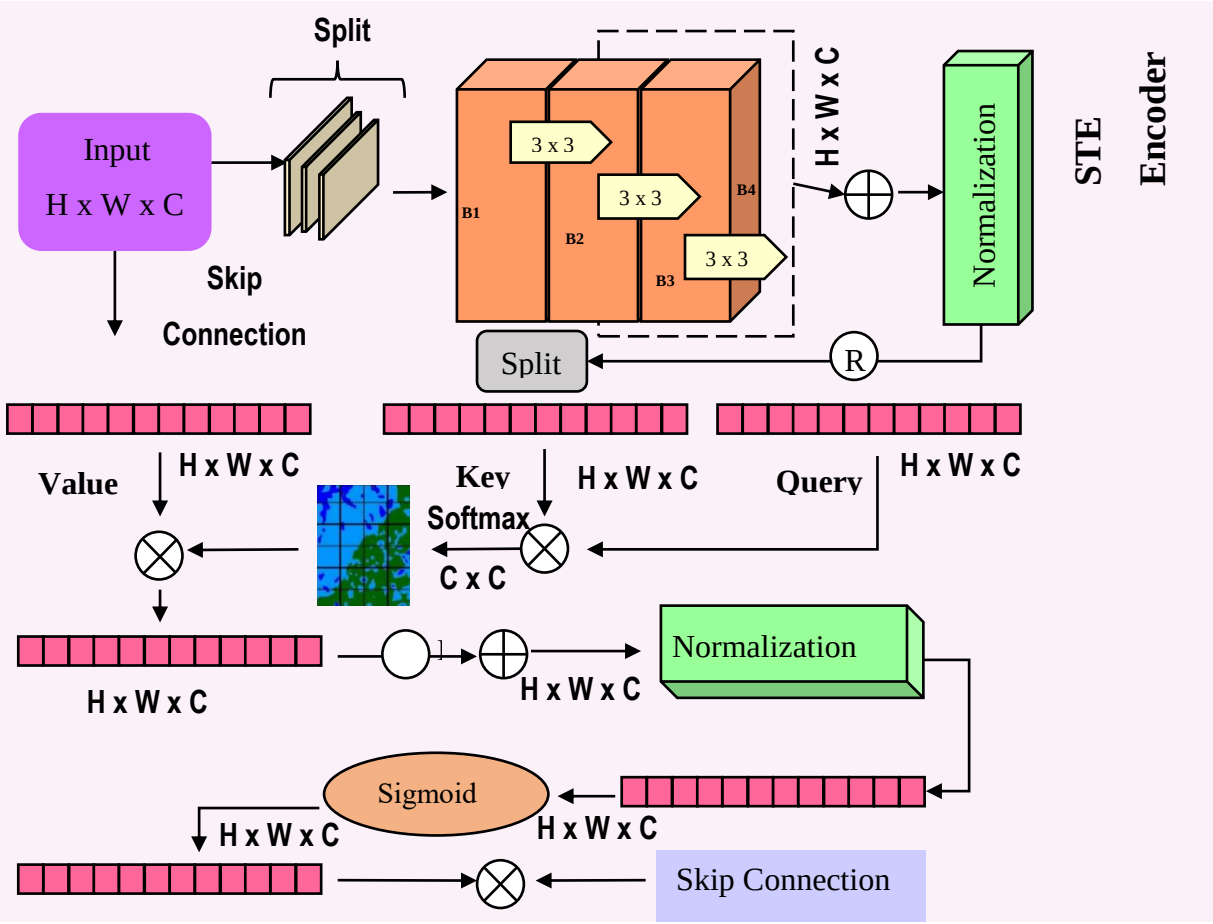


Fig. 5.4 Spatial Transpose Encoder

The STE consists of two main components specifically developed to enhance feature representation and improve understanding of the global context. It is represented in Fig. 5.4. It utilizes cutting-edge methods to improve flexibility and effectiveness in managing various levels of detail and worldwide picture representations.

The first component of the STE focuses on learning adaptive multiscale feature representations. The STE adopts a multi-scale processing approach without using 1×1 pointwise convolution layers, ensuring a lightweight network with optimized parameters and operational efficiency.

The method involves splitting the input tensor of size $\mathbb{H} \times \mathbb{W} \times \mathbb{C}$ into $eight$ subsets, each represented by a_i, the spatial dimensions remain the same, while the number of channels is reduced to $\frac{\mathbb{C}}{\mathbb{s}}$. The $i \in \{1, 2, 3, 4\ldots, \mathbb{s}\}$ and $\mathbb{H}, \mathbb{W}, \mathbb{C}$ denote height, width and channel respectively. Every subset, excluding the initial one, performs a 3x3 channel-wise convolution f_i , generating an outcome denoted as $\mathcal{Z}_i$. In addition, the result of the previous depth-wise convolution, denoted as $\mathcal{Z}_{i-1}$, is combined with the current subset a_i before being processed by f_i. Each depth-wise operation f_i processes feature maps from all preceding splits $a_j, j \leq i$

The cardinality of the set s is dynamically modified according to the stage t, where t belongs to the set { 2, 3, 4}. The output variable $\mathcal{Z}_i$ is defined as shown in eq (5.2):

$$\mathcal{Z}_i = \begin{cases} a_i & if\ i = 1; \\ f_i(a_i) & if\ i = 1;, \mathrm{t} = 2\ ; \\ f_i(a_i) & if\ \ 2 < i \leq s\ and\ t. \end{cases} \tag{5.2}$$

To represent the global context more efficiently, we have implemented a transposed query and critical attention method, which avoids the excessive computational burden associated with typical transformer self-attention layers. This method simplifies the process by calculating the dot-product operation of the multi-head self-attention (MSA) across channel dimensions instead of spatial dimensions. This allows for the computation of cross-covariance across channels, resulting in attention feature maps with a natural awareness of global context. Given a tensor $\mathcal{Z}$ that has been normalized and has a shape of $\mathbb{H} \times \mathbb{W} \times \mathbb{C}$, it computes the projections of the query ($\mathcal{Q}$), key ($\mathcal{K}$), and value ($\mathcal{V}$) using three linear layers as it shows in eq(5.3):

$$\mathcal{Q} = \mathcal{W}_{\mathcal{Q}}\, \mathcal{Z},\ \ \mathcal{K} = \mathcal{W}_{\mathcal{K}}\, \mathcal{Z},\ \ \mathcal{V} = \mathcal{W}_{\mathcal{V}}\, \mathcal{Z} \tag{5.3}$$

The variables $\mathcal{W}_{\mathcal{Q}}$, $\mathcal{W}_{\mathcal{K}}$, and $\mathcal{W}_{\mathcal{V}}$ represent the projection weights for $\mathcal{Q}$, $\mathcal{K}$, and $\mathcal{V}$ respectively. The dimensions of each object are defined by the variables $\mathbb{H}, \mathbb{W}\ and\ \mathbb{C}$. L_2 normalization is used on $\mathcal{Q}$ and $\mathcal{K}$ to provide stability throughout training. Instead of calculating the dot-product between $\mathcal{Q}$ and $\mathcal{K}^{\mathcal{T}}$across the spatial dimension $(\mathbb{H} \cdot \mathbb{W} \times \mathbb{C}) \cdot (\mathbb{C} \times \mathbb{H} \cdot \mathbb{W})$, it calculates across the channel dimensions between $\mathcal{Q}^{\mathcal{T}}$and $\mathcal{K}$ $(\mathbb{C} \times \mathbb{H} \cdot \mathbb{W}) \cdot (\mathbb{H} \cdot \mathbb{W} \times \mathbb{C})$, resulting in a C×C softmax-scaled attention score matrix. The final attention maps are derived by multiplying the scores with the matrix $\mathcal{V}$ and then summing them. Afterward, two 1x1 pointwise convolution layers, layer normalization (LN) and GELU activation generate non-linear features as it is demonstrated in eq (5.4) & (5.5).

$$\widehat{\mathcal{X}} = TransposeAttention(\mathcal{Q}, \mathcal{K}, \mathcal{V}) + \mathcal{X} \tag{5.4}$$

$$TransposeAttention(\mathcal{Q}, \mathcal{K}, \mathcal{V}) = \mathcal{V} \cdot \text{softmax}(\mathcal{Q}^{\mathcal{T}} \cdot \mathcal{K}) \tag{5.5}$$

Hierarchical Transpose Attention (HTA)

The Hierarchical Transpose Attention (HTA) architecture combines components of Vision Transformers (ViTs) and Convolutional Neural Networks (CNNs) to produce accurate and efficient feature representation, as can be observed in Fig. 5.5. The architecture begins by analyzing an input image with dimensions of 224×224. The image is then separated into smaller patches, which can be 14×14 or 7×7. The patches are converted into linear embedding (tokens) using a standard linear layer, following the approach suggested by Tolstikhin et al. (2021)[75]. The tokens are then inputted into a sequence of HTA blocks specifically intended to encode spatial and channel information. Once the HTA blocks have been processed, the tokens are averaged across the spatial dimensions and fed into a fully connected layer to provide the final class predictions.

The HTA block, which serves as the foundational unit of this architecture, has two primary modules: Permute-MLP and Channel-MLP. The Permute-MLP module stores spatial information, while the Channel-MLP module encodes channel information. The Channel-MLP module is organized comparably to the feed-forward layer in Transformers[74], consisting of two wholly linked layers with a GELU activation function in the middle.

The Permute-MLP in the HTA block performs a distinct operation on three-dimensional token representations by utilizing three branches. Each branch encodes information specifically along

the height, breadth, or channel dimension. The process of spatial information encoding consists of a height-channel permutation operation, which is then followed by a fully linked layer to combine and include the spatial information. Given an input tensor $\mathbb{X}$ that belongs to the actual numbers and has dimensions as shown in eq(5.6-5.10).

$$\mathbb{H} \times \mathbb{W} \times \mathbb{C}. \tag{5.6}$$

$$\mathbb{X}_{\mathbb{H}} = \mathbb{P}\mathtt{ermute} - \mathbb{H}(\mathbb{X}), \tag{5.7}$$

$$\mathbb{X}_{\mathbb{W}} = \mathbb{P}\mathtt{ermute} - \mathbb{W}(\mathbb{X}), \tag{5.8}$$

$$\mathbb{X}_{\mathbb{C}} = \mathbb{F}\mathtt{ully}\ \mathbb{C}\mathtt{onnected}_{\mathbb{C}}(\mathbb{X}) \tag{5.9}$$

$$\widehat{\mathbb{X}} = \mathbb{F}\mathtt{ully}\ \mathbb{C}\mathtt{onnected}(\mathbb{X}_{\mathbb{H}}, \mathbb{X}_{\mathbb{W}}, \mathbb{X}_{\mathbb{C}}), \tag{5.10}$$

To improve the Permute-MLP, we propose the Weighted Permute-MLP, which adjusts the significance of various branches dynamically employing divided attention. This enhanced approach prioritizes optimizing the network's ability to emphasize the most relevant aspects, namely focusing on $\mathbb{X}_{\mathbb{H}}, \mathbb{X}_{\mathbb{W}}$ & $\mathbb{X}_{\mathbb{C}}$. The HTA network consists of many HTA blocks, each carrying out adaptive multiscale feature learning and efficient global context encoding. The partnership of

the Permute-MLP and Channel-MLP modules effectively captures complex spatial and channel interactions, guaranteeing a robust and efficient feature representation.

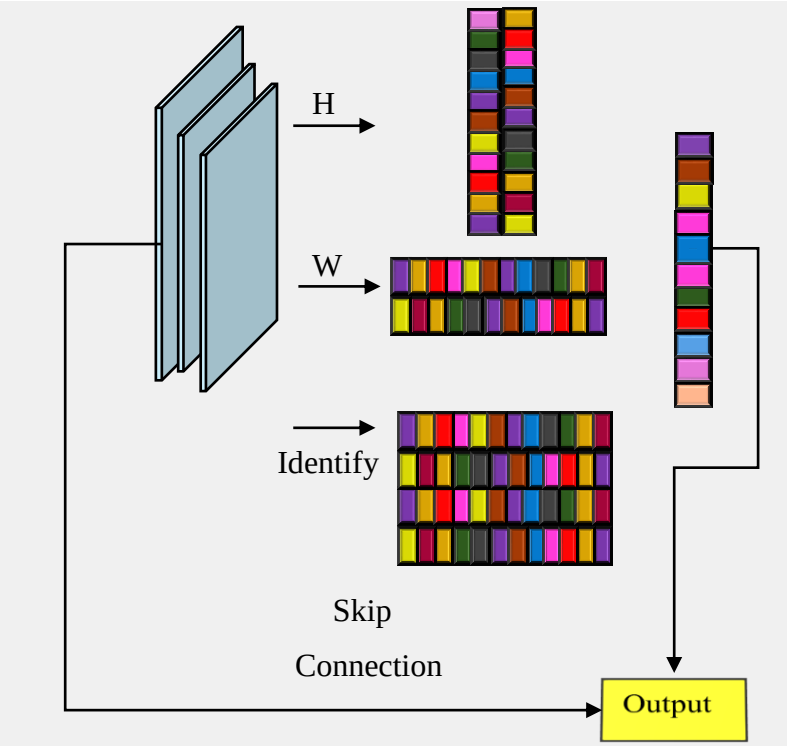


Fig. 5.5 Hierarchical Transpose Attention

In summary, the HTA architecture is a notable improvement that effectively manages the complexity of models and computing efficiency. The architecture of this system enables efficient processing and achieves high accuracy, effectively tackling the specific issues presented by surroundings with limited resources.

5.2.3 Experimental Results and Discussion

This section contains detailed information regarding the dataset utilized during the research, the experimental settings of the proposed framework, and performance assessments.

Dataset and preprocessing

To evaluate the effectiveness of the Arctic-Net model, we employed an extensive SAR dataset consisting of 4,000 sea ice pictures that were categorized into seven unique groups. Every image was adjusted to a resolution of 224×224 pixels to maintain uniformity throughout the collection.

The dataset was divided into three subsets: training (70%), validation (15%), and testing (15%). Careful attention was given to ensuring that each division maintained a balanced distribution of classes.

Normalization was a crucial step in preprocessing, including adjusting pixel values to have a mean of zero and a standard deviation of one based on the dataset. Constraining the pixel values to a consistent range improved the efficiency of training the neural network. After normalizing, the dataset was divided into 80% for training, 10% for validation, and 10% for testing. Stratified sampling was used to equally represent all sea ice categories in these subgroups.

The data loading process was overseen by PyTorch's DataLoader, which was set up to handle batch sizes and shuffling optimally, specifically designed for the training, validation, and test sets. This configuration was explicitly created to accelerate importing and grouping data during the training stage. In addition, data augmentation techniques were utilized, such as random horizontal and vertical flips, to improve the model's capacity to generalize. The preparation process, which includes data augmentation, normalization, and efficient data loading, greatly enhanced the quality and variety of the training dataset. As a result, the performance and generalization of the Arctic-Net model in the sea ice classification task were dramatically improved.

Implementation details

The PyTorch implementation of the Arctic-Net model consists of three main components: the Adaptive Convolutional Encoder (ACE), the Spatial Transposer Encoder (STE), and the Hierarchical Transpose Attention (HTA) mechanism. Every element is carefully crafted to maximize computing efficiency while maintaining the integrity of feature representation.

- **Adaptive Convolutional Encoder (ACE)**: ACE is built utilizing depth-wise separable convolutions, which use dynamic kernel sizes (3, 5, 7, and 9). This design enables the effective extraction of local features in a resource-efficient manner. Layer Normalization (LN) and Gaussian Error Linear Unit (GELU) activations are used to map non-linear features efficiently. ACE incorporates skip connections to enable continuous information propagation inside the network.
- **Spatial Transposer Encoder (STE)**: STE is designed to acquire flexible and comprehensive feature representations at several scales and improve the encoding of

overall context. The input tensor is partitioned into eight subgroups, and each subset undergoes depth-wise convolutions. An efficient method is used to capture global context by employing a transposed query and key attention mechanism, which avoids the computing burden usually associated with traditional self-attention layers.

- **Hierarchical Transpose Attention (HTA)**: The HTA model combines features from Vision Transformers (ViTs) and Convolutional Neural Networks (CNNs) to capture both spatial and channel information. The architecture consists of a sequence of HTA blocks, each consisting of Permute-MLP and Channel-MLP modules. The Permute-MLP module specifically emphasizes the encoding of spatial data, whereas the Channel-MLP module is responsible for processing channel information. The addition of weighted permute-MLP incorporates dynamic attention, improving the model's capacity to prioritize essential characteristics.

Training Protocols

The Arctic-Net model underwent training using the AdamW optimizer, with an initial learning rate of 0.001. A cosine annealing learning rate schedule was implemented to reduce the learning rate, following a cosine curve progressively. This approach was shown to be successful in improving the learning process in the later phases of training. The training was carried out for 100 epochs using a batch size of 32. The model's generalization capabilities were enhanced by applying data augmentation techniques, such as random rotations, horizontal and vertical flips, and normalization.

Hardware

The studies utilized a machine with two NVIDIA A5000 GPUs, each possessing 24 GB of RAM, to enhance the deep learning calculations for Arctic-Net model training. The device was also equipped with 128 GB of RAM and SSD storage to effectively manage the extensive SAR dataset and guarantee seamless data processing. The model was implemented and optimized using PyTorch's deep learning framework, utilizing GPU acceleration via CUDA for enhanced speed.

Performance Metrics

The performance of the trained model is assessed using a range of evaluation metrics, including accuracy, precision, recall, F1-score, and Matthews Correlation Coefficient (MCC). These metrics offer comprehensive insights into various facets of the model's classification performance.

Accuracy: It measures the accuracy of the model's predictions by calculating the ratio of correctly classified samples to the total number of samples in the dataset. However, accuracy may provide a partial picture, especially in class imbalance.

Precision: This metric focuses on the correctness of optimistic predictions and measures the proportion of accurate positive predictions among all optimistic predictions made by the model. It is beneficial when the cost of false positives is high.

Recall: Also known as sensitivity or actual positive rate, recall measures the proportion of accurate optimistic predictions among all actual positive samples in the dataset. It evaluates the model's ability to capture all relevant instances of a particular class.

F1-score: The F1-score is the harmonic mean of precision and recall, providing a balanced measure of the model's performance. It is precious when there is an imbalance between the number of positive and negative instances in the dataset.

Matthews Correlation Coefficient (MCC): MCC is a correlation coefficient that considers true and false positives and negatives. It ranges from -1 to 1, with 1 indicating perfect predictions, 0 indicating random predictions, and -1 indicating complete disagreement between predictions and ground truth.

Experimental Results and Analysis

This section evaluates the proposed Arctic-Net model compared to many advanced models for sea ice classification, including PGN+SVM, PGIL, Vision Transformer (ViT), DenseNet, DaVIT, ResNext, and Swin Transformer. This study utilizes Precision, Recall, F1 Score, and Accuracy as evaluation measures to examine each model's performance thoroughly.

Table 5.2 Performance Metrics and Comparison Again SOTA

Model	Performance Metrics			
	Precision	**Recall**	**F1 score**	**Accuracy**

PGN+SVM[56]	0.83	0.82	0.82	0.84
PGIL[56]	0.85	0.85	0.84	0.86
VIT[76]	0.85	0.86	0.85	0.87
DenseNet[77]	0.89	0.88	0.87	0.89
DaVIT[78]	0.84	0.84	0.84	0.85
ResNext[79]	0.85	0.85	0.85	0.86
Swin[80]	0.87	0.87	0.86	0.87
Proposed(Arctic-Net)	**0.91**	**0.92**	**0.91**	**0.93**

The findings in Table 5.2 indicate that Arctic-Net(highlighted in bold) surpasses all rival models in the assessed criteria, highlighting its exceptional categorization proficiency.

The Arctic-Net model has superior performance across all parameters, with a precision of 0.92, recall of 0.92, F1 score of 0.91, and accuracy of 0.93. These results indicate a substantial advancement compared to previous models, underscoring Arctic-Net's superior capacity to reliably categorize various sea ice forms.

The enhancements in recall and F1 score illustrate Arctic-Net's capability to accurately detect pertinent events across all categories, particularly infrequent or more challenging to classify. The equilibrium between accuracy and recall is essential in practical contexts like sea ice monitoring, where both overestimating and underestimating particular ice types can significantly affect marine navigation and climate research.

Compared to the baseline models and prior research, including PGN+SVM and PGIL[56]. Arctic-Net significantly enhances accuracy and recall, signifying a more effective equilibrium between reliably recognizing positive samples and minimizing false positives. This benefit is more apparent when juxtaposing Arctic-Net with more sophisticated designs like ViT, DenseNet, ResNext, and Swin. While DenseNet attains a commendable accuracy of 0.89, Arctic-Net outperforms all SOTA models, underscoring the superiority of its architectural elements. The improvements in Arctic-Net's performance are due to its distinctive architecture, which incorporates the Adaptive Convolutional Encoder (ACE), Spatial Transposer Encoder (STE), and

Hierarchical Transpose Attention (HTA) methods. This combination efficiently collects local and global information, optimizing accuracy and model complexity.

The experimental findings validate the efficacy of Arctic-Net as a reliable approach for sea ice categorization. The model's capacity to attain elevated accuracy while maintaining balanced precision and recall renders it especially appropriate for practical applications where computing resources are constrained, yet superior classification performance is essential.

Qualitative analysis of Arctic-Net

The t-SNE visualization of Arctic-Net's output offers a lucid representation of the model's ability to distinguish between sea-ice types, which can be observed in Fig. 6. The scatter figure demonstrates that the model proficiently clusters many categories, including Sea-Young Ice and sea-old-ice, signifying excellent differentiation between these classes. Nevertheless, categories with limited samples, such as Isbergs-Glacier and Icebergs, have more varied distributions. This indicates that although the model excels in more common categories, there is room for enhancement in differentiating less frequent classifications. The t-SNE results underscore the model's ability to delineate unique feature representations of several sea-ice kinds while identifying improvement opportunities.

5.2.4 Conclusion

This study presented the Arctic-Net model, an innovative hybrid framework that combines the advantages of Convolutional Neural Networks (CNNs) and attention processes to proficiently categorize sea ice from SAR pictures. Arctic-Net exhibited enhanced accuracy and computing efficiency by utilizing the Adaptive Convolutional Encoder (ACE), Spatial Transposer Encoder (STE), and Hierarchical Transpose Attention (HTA) components. The model surpassed leading methodologies, including Vision Transformers (ViT), DenseNet, and Swin Transformer, demonstrating significant enhancements in precision, recall, F1-score, and overall accuracy.

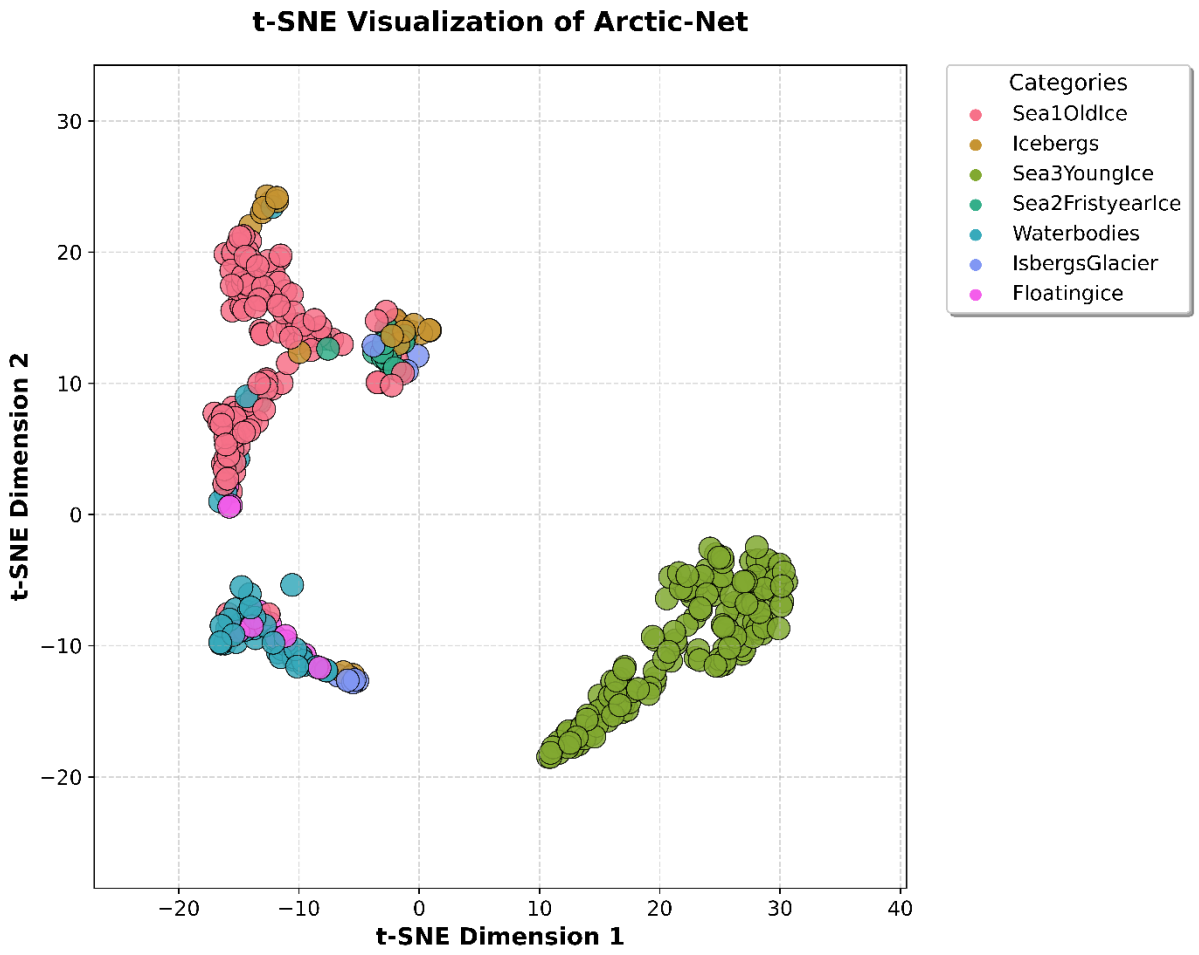


Fig. 5.6 t-SNE visualization of Arctic-Net embedding showing clustering of sea-ice categories.

Arctic-Net can derive local and global contextual information from intricate SAR data, facilitating accurate sea-ice classification, especially in demanding conditions with few labeled data. Its lightweight construction renders it exceptionally appropriate for deployment in resource-limited environments, hence broadening its potential applications in climate monitoring, marine navigation, and environmental research.

Although the model demonstrates considerable progress, subsequent research should investigate improving its scalability and resilience to diverse sensor inputs and seasonal fluctuations. Moreover, subsequent research may integrate more varied datasets and real-time processing functionalities to enhance its applicability in operational sea-ice monitoring systems.

5.3 Significant Outcomes of this Chapter

The significant outcomes of this chapter are as follows:

- To predict fire risk levels using remote sensing imagery through a novel framework named "IGNITE-NET" (Fire Risk Prediction using Dynamic Receptive Fields and Dynamic Channel Fusion Attention). The proposed model comprises two primary modules: Dynamic Receptive Field Blocks (DRFBs) for efficient feature extraction with reduced computational complexity and Dynamic Channel Fusion Attention (DCFA) for optimized cross-channel interactions, enhancing the predictive accuracy without dimensionality reduction.
- Implemented Self-Supervised Knowledge Distillation (SSKD) to improve model generalization and robustness, enabling the use of both annotated and unannotated datasets for enhanced learning outcomes.
- Conducted extensive performance evaluations, including accuracy, precision, recall, F1 score, and Matthews Correlation Coefficient (MCC), demonstrating the model's superior performance over state-of-the-art approaches.
- Performed ablation and generalization studies to assess the resilience and robustness of the proposed IGNITE-NET model across various environmental conditions and datasets, ensuring its applicability in real-world fire risk assessment scenarios.

The following research studies serve as the foundation for this chapter:

- ❖ Abhishek Verma, Virender Ranga, Dinesh Kumar Vishwakarma, "Arctic-Net: A Hybrid Convolutional and Attention-Based Model for Efficient Sea Ice Classification Using SAR Images" Communicated in Cluster Computing
- ❖ Indian patent published ,application number 202511033470 "System and Method for monitoring Sea Ice Formations Fire in Real time

This chapter introduces Arctic-Net, a novel hybrid deep learning framework that integrates CNNs with attention mechanisms to enhance sea ice classification using SAR images, achieving significant performance improvements over existing models.

The next chapter focusses on IGNITE-NET, a cutting-edge fire risk prediction model that combines Dynamic Receptive Field Blocks (DRFBs) and Dynamic Channel Fusion Attention (DCFA) to deliver accurate and efficient fire risk assessments, marking an important step forward in disaster management and environmental monitoring.

Chapter 6: IGNITE-NET: DYNAMIC ATTENTION DRIVEN FIRE RISK PREDICTION

6.1 Scope of this chapter

Accurate fire risk prediction has become a pivotal element in environmental conservation, disaster management, and public safety in the context of escalating wildfire occurrences due to climate change and urban expansion. The ability to efficiently assess fire risks using remote sensing data and Advanced Machine Learning and Deep Learning Techniques is crucial for mitigating the adverse impacts of wildfires on human life, infrastructure, and ecosystems. Traditional fire risk assessment models, including physics-based simulations and classical statistical methods, often grapple with high computational demands and limited feature extraction capabilities, which restrict their effectiveness and scalability.

To address these challenges, this chapter introduces IGNITE-NET, an innovative deep learning framework tailored for fire risk prediction. IGNITE-NET integrates Dynamic Receptive Field Blocks (DRFBs) and Dynamic Channel Fusion Attention (DCFA) mechanisms within a lightweight Convolutional Neural Network (CNN) architecture. This integration leverages dynamic feature extraction and attention-based optimization strengths to enhance local cross-channel interactions and maintain high-dimensional feature integrity. The proposed methodology significantly reduces computational complexity while achieving superior predictive accuracy, outperforming existing models such as HRNET, ResNext, and Max ViT.

This chapter provides a detailed exposition of the IGNITE-NET framework, elucidating its core components: DRFBs, which optimize spatial and channel-wise feature interactions, and DCFA, which refines channel attention predictions without dimensionality reduction. The FireRisk dataset, derived from the Wildfire Hazard Potential (WHP) dataset and the National Agriculture Imagery Program (NAIP), serves as the primary data source for training and evaluation. The dataset section covers pre-processing techniques including image normalization, augmentation, and stratified sampling to enhance model robustness and generalization. The experimental setup is comprehensively discussed, detailing the hardware configurations, software environment, and hyper parameter tuning strategies employed to optimize model performance. Performance metrics

such as Accuracy, Precision, Recall, F1 Score, and Matthews Correlation Coefficient (MCC) are used to rigorously evaluate IGNITE-NET, with comparative analyses against state-of-the-art models demonstrating its robustness and reliability. Additionally, this chapter highlights qualitative analyses using visualization techniques like t-SNE and Layer CAM, which provide insights into the model's decision-making process and its ability to discriminate between different fire risk levels. The integration of Self-Supervised Knowledge Distillation (SSKD) is also explored, showcasing its role in enhancing model generalization and reducing overfitting.

By introducing IGNITE-NET, this research advances the field of fire risk assessment, offering scalable and efficient solutions for environmental monitoring and wildfire management. The findings contribute to the development of proactive fire mitigation strategies, informing policy decisions and supporting real-time fire risk monitoring systems. Future research directions include the incorporation of additional geospatial data sources, the application of advanced data augmentation techniques, and the exploration of real-time deployment scenarios to further enhance the practical applicability of IGNITE-NET in diverse environmental settings.

6.2 IGNITE-NET: Fire Risk Prediction using Dynamic Receptive Fields and Dynamic Channel Fusion Attention

6.2.1 Abstract

Forecasting the likelihood of fires is crucial for reducing the severe impacts of wildfires, making it a key component of environmental conservation and public protection. Identifying fire-prone areas promptly and accurately allows for taking pre-emptive steps, minimizing risks to human lives, property, and ecosystems. Current approaches to predicting fire danger struggle with computational complexity and inefficiency in extracting features. This research presents IGNITE-NET, a method for assessing fire risk that utilizes advanced deep neural network topologies and attention mechanisms. The IGNITE-NET system comprises two main components: Dynamic Receptive Field Blocks (DRFBs) and Dynamic Channel Fusion Attention (DCFA). The components improve existing approaches by reducing computational costs, retaining feature quality, and capturing local cross-channel interactions without reducing dimensionality. IGNITE-NET also utilizes Self-Supervised Knowledge Distillation (SSKD) to improve the model's performance and generalization skills. Experimental evaluations show that IGNITE-NET

outperforms existing models in crucial performance measures like test accuracy, Matthews Correlation Coefficient, precision, recall, and F1 score. Collaboration is invited to improve the feasibility and application of the suggested model in real-world situations. IGNITE-NET is a significant step forward in fire risk assessment, providing creative ways to tackle persistent difficulties and support proactive wildfire control tactics.

6.2.2 Proposed Methodology

The proposed methodology will be discussed in this section.

Dynamic Receptive Field Blocks (DRFBs)

The ResNet architecture has become a reliable foundation for addressing diverse computer vision tasks [35], with skip connections effectively mitigating vanishing gradient issues in deeper models. Inspired by ResNeSt[79]: Split-Attention Networks, which employs multi-branch channel-wise attention to enhance representation learning, we propose Dynamic Receptive Field Blocks (DRFBs). DRFBs extend the Split-Attention mechanism by dynamically adjusting receptive fields, enabling enhanced spatial and channel-wise feature interactions. This design improves the network's ability to capture complex spatial details, making it ideal for fire risk assessment, as given by Equation (6.1).

$$i_{output} = \mathfrak{B}_{r}(\mathcal{A}_{u}(\mathcal{L}^{Conv_3}(\mathcal{C}d(input)))) + i_{input} \tag{6.1}$$

The input feature map is represented as $input \in \mathbb{R}^{C\times H\times W}$. The notation $\mathcal{C}d(.)$ denotes a 1×1 convolutional operation that decreases the number of channels (C) in the input feature map to a new number of channels C', where $C > C'$. $\mathfrak{B}_{r}(.)$ is a concatenation of batch normalization and rectified linear unit (ReLU) activation. The operation $\mathcal{L}^{Conv_3}(.)$ indicates a 3×3 convolutional operation that captures important non-linear correlations. This operation maintains the same number of feature channels while gradually reducing the spatial dimension, which refers to the height and width of the feature maps. $\mathcal{A}_{u}(.)$ denotes a 1×1 convolution that increases the resolution of features with C' channels to C.

In this work, a novel "Dynamic Receptive Field Blocks" (DRFB) module is designed to reduce the computational cost of feature extraction without compromising the quality of features. Specifically, the single-branch feature extraction of Eq. 1 can be replaced by a multi-branch design

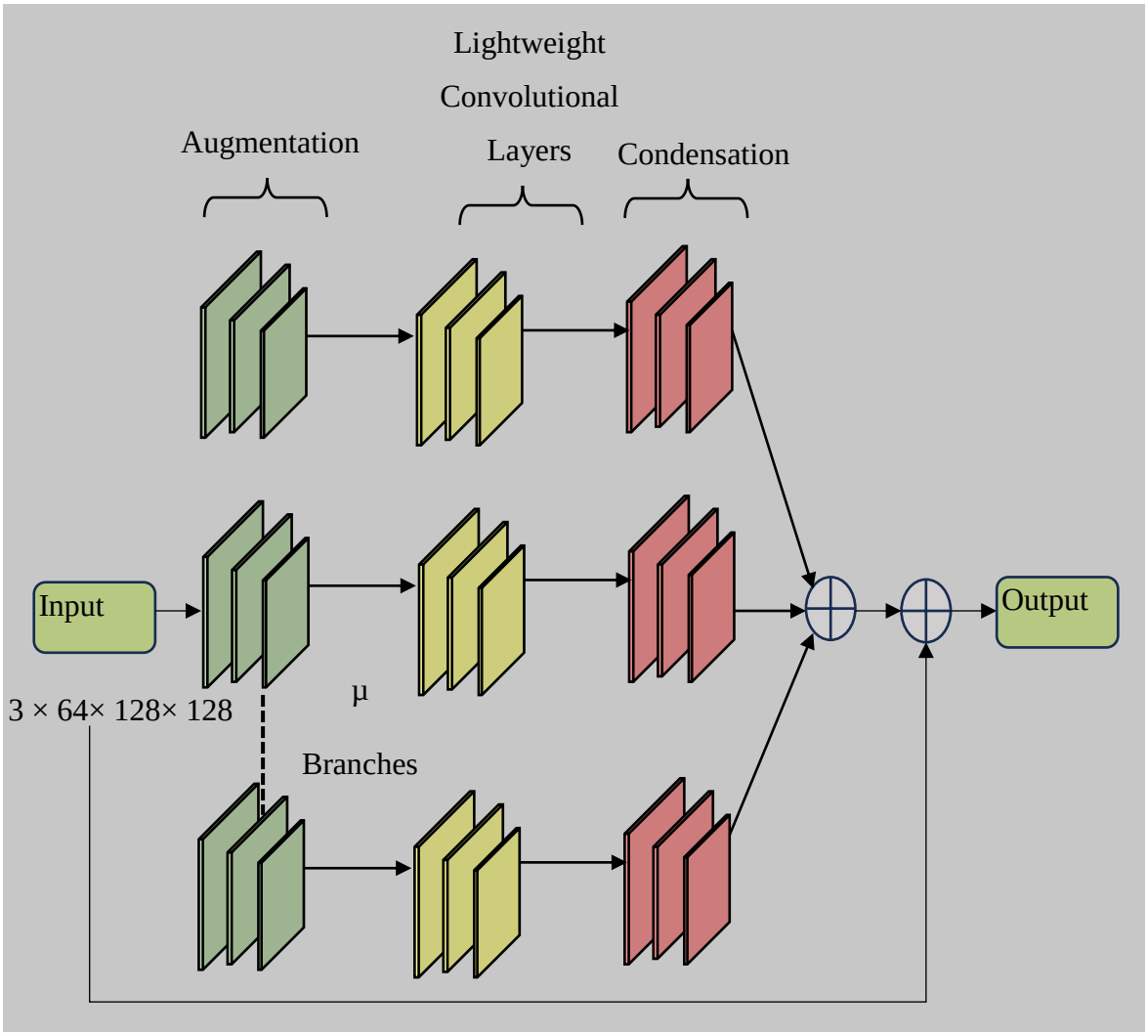


Fig. 6.1 The proposed Dynamic Receptive Field Blocks (DRFBs) module

that reduces the computational overheads. Specifically, the number of sub-branches is defined by a hyperparameter μ such that the $\mathcal{L}^{Conv_3}(.)$ operation reduces each feature map to $C'' = 2 \times {}^{C}/_{\mu}$ channels in each sub-branch, thereby significantly reducing the computational complexity. It is noteworthy that $C > C' > C''$.The resulting feature maps from each sub-branch after the squeeze, convolution, and unsqueeze operations are summed elementwise, which is represented by the $\sum(.)$ in Eq. 2. Here $\mathcal{A}_{u\,C''}(.)$, $\mathcal{L}^{Conv_3}{}_{C''}(.)$ and $\mathcal{C}d_{C''}(.)$ represent similar operations from Eq. 6.1 with reduced computation.

$$i_{output} = \sum_{k=1}^{\mu}(\mathfrak{B}_r(\mathcal{A}_{u\,C''}(\mathcal{L}^{Conv_3}{}_{C''}(\mathcal{C}d_{C''}(i_{input}))))) + i_{input} \tag{6.2}$$

For example, let the input be $i_{input} \in \mathbb{R}^{64\times32\times32}$ which is the input feature size of the first skip connection block of the Fig. 6.2 Dynamic Receptive Field Blocks module ResNet architecture. In

the ResNet architecture, the 3×3 convolution operation produces 64 channel features of the same spatial dimension. The total number of parameters in this lightweight convolutional layer is 36928. However, constructing the proposed DRFB module with $\mu = 32$, the 3×3 convolution $Conv_{3_{C''}}(.)$ produces 4 channel output from 4-channel input. It is noteworthy that the 1×1 convolution $\mathcal{C}\mathcal{d}_{C''}(.)$ in the DRFB downsamples 64 channel inputs to 4 channels for each sub-branch before passing to $\mathcal{L}^{Conv_3}{}_{C''}(.)$. Each convolution layer in this design contains merely 148 parameters. The total number of convolutional parameters across all the sub-branches is $148 \times 32 = 4736$, significantly less than the convolutional block of the standard ResNet. Similarly, for $\mu = 64$, the proposed DRFB module has a convolutional layer with just 2432 parameters, clearly demonstrating the lightweight nature of the proposed DRFB blocks against each of the four ResNet skip connection blocks. Proposing such a multi-branch architecture helps achieve superior performance at reduced computation without increasing the depth of the neural network, which often leads to overfitting [81].

The DRFB feature extractor can simulate the representational power typically achieved by larger and denser layers without incurring the computational complexity that is commonly associated with such layers. This is accomplished by splitting the input into lower-dimensional embedding using unit convolutions, followed by transformation using the same set of filters in parallel branches. Finally, the transformed embedding is concatenated to achieve the desired consolidated transformation. The uniform across all multiple branches has significant implications for model complexity, as it minimizes the need for fine-tuning a large number of hyper parameters that would have been required if each branch had a distinct design.

Dynamic Channel Fusion Attention (DCFA)

This section introduces Dynamic Channel Fusion Attention (DCFA), inspired by DynaMixer: A Vision MLP Architecture with Dynamic Mixing[82]. While DynaMixer employs dynamic token-wise fusion to enhance MLP-like models, DCFA adapts this concept for channel interactions. Unlike traditional methods like the SE block, which use dimensionality reduction, DCFA efficiently captures local cross-channel interactions without reducing dimensionality. Building on DynaMixer's dynamic fusion principles, DCFA enhances channel attention prediction, ensuring robust performance for tasks like fire risk assessment..

The operation of DCFA can be broken down into three discrete stages, which also can be observed in Fig. 6.1.

Global Feature Representation: To begin, we independently apply global average pooling to each channel of the input feature maps. Input feature map X of size H × W × C, where H is the height, W is the width, and C is the number of channels. This process extracts a 1 × 1 × C feature vector, where C denotes the number of channels. Implement global average pooling on each channel separately to generate a feature vector.

$$F_{avg}(c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X(i, j, c) \tag{6.3}$$

Importance Estimation: The feature vector's importance is estimated by using a one-dimensional convolution with a kernel size of 1 × k, where k is computed dynamically dependent on channel dimension C.

$$F_{conv} = Conv1D\left(F_{avg}, W_k\right) + b \tag{6.4}$$

$$F_{imp} = ReLU(F_{conv}) \tag{6.5}$$

Regularization using Activation Function: The prediction of importance is constrained between 0 and 1 through the application of an appropriate activation function, such as the sigmoid function.

$$F_{sig} = Sigmoid(F_{imp}) \tag{6.6}$$

This section introduces a new method for creating a specialized deep neural network for detecting FireRisk levels, focusing on optimizing both cost and accuracy. We are implementing the DCFA module to improve the network's learning skills. We use a lightweight CNN framework that incorporates multi-scale feature fusion to perform risk analysis successfully. We explain how the SSKD technique is used to train the network model, guaranteeing good performance effectively.

The SE block, commonly used in several models, typically uses global average pooling, two fully linked layers with non-linearity, and ends with a sigmoid function to produce channel weights. Although this method successfully captures cross-channel interaction and reduces dimensionality to handle model complexity, recent research has shown that dimensionality

reduction has a negative effect on channel attention prediction, making it inefficient and unnecessary to capture dependencies across all channels.

Observed in Fig 3 for a graphic depiction of our module. This module comprises three primary steps: first, performing global average pooling on feature maps to produce a 1 × 1 × C feature vector; second, evaluating the significance prediction of the feature vector via one-dimensional convolution with a kernel size of 1×k; and finally, normalizing the significance prediction to a range of 0 to 1 using the Sigmoid function. Refer to Fig 3 for a graphical representation of our module.

Dataset Description

The FireRisk dataset, referenced as[83], is a carefully curated compilation of remote-sensing images that have been rigorously organized to assess the danger of fire. FireRisk is a crucial resource in remote sensing and environmental monitoring. FireRisk will be compared with several cutting-edge prediction models for fire risk categorization.

The United States Department of Agriculture's Wildfire Hazard Potential (WHP) study is the primary source from which the FireRisk dataset was generated. The comprehensive analysis of fire-risk danger and wildfire severity in different settings has influenced the research community. The 2020 edition of the WHP raster dataset gives a thorough and complete evaluation of fire-risk hazards in different parts of the US. Many geostatistical datasets were used to compile this assessment. For instance, the Fire Programme Analysis (FPA) was used to compile a dataset on the frequency and severity of fires, Wildfire was used to compile data on fuels and vegetation, and FSim was used to estimate the likelihood and severity of wildfires.

The FireRisk dataset encompasses a full knowledge of the complex dynamics and nuanced aspects of fire risk assessment, by including various and multiple information from the WHP project. The FireRisk dataset plays a crucial role in our research by providing a robust and reliable foundation for assessing and refining fire risk prediction algorithms. Its comprehensive structure and carefully curated data support a systematic evaluation process, enabling accurate benchmarking and improvement of predictive methodologies.

To effectively assess fire risk, the FireRisk dataset serves as an essential resource, offering comprehensive and detailed information. By analyzing its spatial and temporal data, we can

identify and explain critical patterns and trends influencing fire risk dynamics. This enables a deeper understanding of the factors driving fire risk variability across diverse environments and timeframes.

This research shows that the FireRisk dataset is an important source for building and testing algorithms to anticipate fire risks. By utilizing the extensive information in this carefully selected dataset, we can accurately assess the effectiveness and reliability of different prediction.

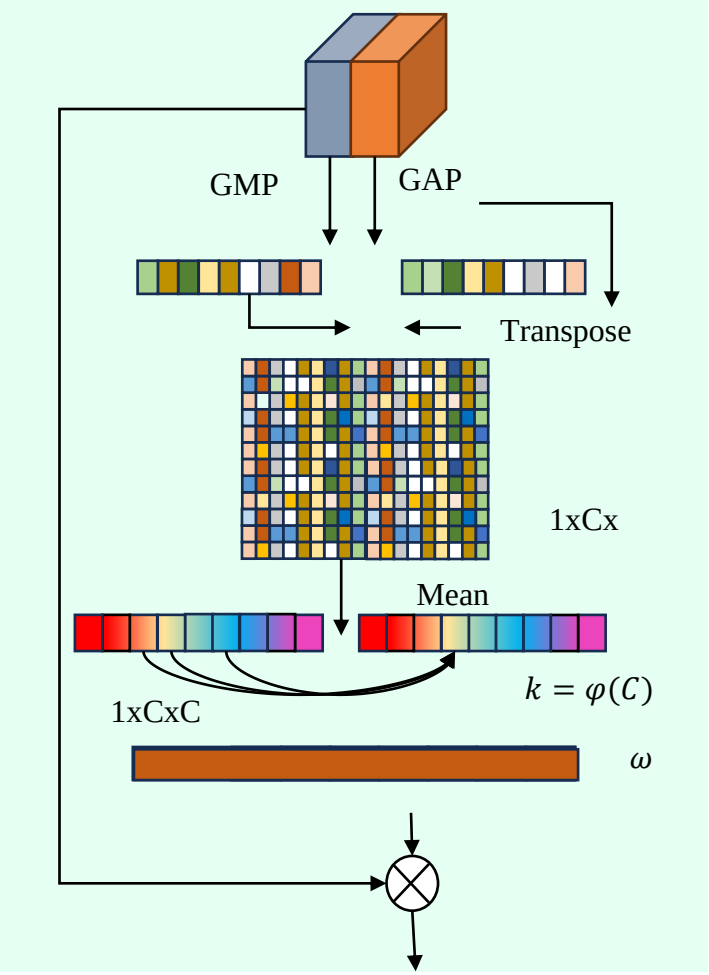


Fig. 6.3 Diagram of Dynamic Channel Fusion Attention (DCFA).

Table 6.1 Pseudocode of the Proposed " IGNITE-NET " Model

Input: **Dataset=** $\{X_i, Y_i\}_{i=1}^{N}$, $X_i \in R^{3\times270\times270}$ illustrating input images & $Y_i \in$**{0,1,2,3,4,5,6}** as Corresponding labels

Model Parameters $\boldsymbol{\theta}$

Batch Size $\boldsymbol{B}$

Epoch $\boldsymbol{\varepsilon}$

Learning Rate $\boldsymbol{l_r}$

After $\boldsymbol{n}$ epochs, Learning Rate Decay Factor $\boldsymbol{\gamma}$, where $\boldsymbol{\gamma} \in [\mathbf{0}, \mathbf{1}]$

Output: Trained IGNITE-NET model for Fire Risk assessment

Initialize $\boldsymbol{\theta}$ and weights $\boldsymbol{\alpha}$

for $\boldsymbol{i} = \mathbf{1}$ ... ... $\boldsymbol{\varepsilon}$ **do** (Train for a certain $\boldsymbol{\varepsilon}$ number of epochs)

for $\boldsymbol{j} = \boldsymbol{x_1}$ ... ... $\boldsymbol{x_B}$ **do** (Iterate through each batch $\boldsymbol{B}$ within the

$(\boldsymbol{x_B}, \boldsymbol{y_B})\boldsymbol{\rho}$ (Randomly select one batch with a size of $\boldsymbol{B}$)

$\widehat{\boldsymbol{y}}_{\boldsymbol{b}} = (\boldsymbol{x_B}; \boldsymbol{\theta})$ (Compute posterior probability for each input sequence)

$\boldsymbol{L_{CE}}$=$\boldsymbol{cross\ entropy\ loss}(\boldsymbol{y}, \boldsymbol{y_b})$ (Calculate cross-entropy loss)

$\boldsymbol{\theta} \Leftarrow \boldsymbol{\theta} - \boldsymbol{l_r}\boldsymbol{\Delta_\theta}\boldsymbol{L_{CE}}(\boldsymbol{y_F}, \boldsymbol{y})$ (Optimize model parameters by minimizing.

computing loss{ $\boldsymbol{\theta}$ } cross-entropy loss $\boldsymbol{L_{CE}}$ using backpropagation)

if $(\boldsymbol{i}\%\boldsymbol{n} == \mathbf{0})$

end for

end for

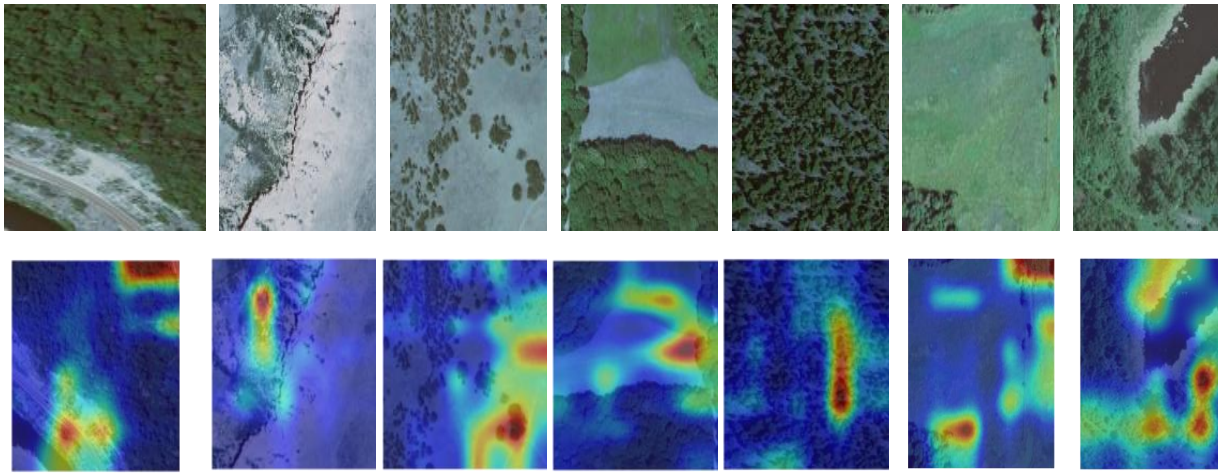
Return None

Integrating with the WHP project and drawing from an array of geo-statistical resources, the Fire-risk dataset is also very detailed. The forecasting models are more accurate and reliable as a result.

To conduct comprehensive investigations and develop robust models for predicting fire risks, the FireRisk dataset is an integral aspect of our research infrastructure. As a comprehensive tool for evaluating fire risk, the WHP project distinguishes itself because it integrates many geostatistical data sources. More effective approaches have been developed for reducing wildfires based on this increased understanding of wildfire dynamics.

Overall Model Structure

The fire risk assessment model name IGNITE-NET that incorporates two main components: the "Dynamic Receptive Field Blocks" (DRFBs) and the "Dynamic Channel Fusion Attention" (DCFA) module it is demonstrated in Fig. 6.5, which aim to improve feature extraction and efficiently record local cross-channel interactions. These components are essential for creating a strong deep neural network designed to identify fire risk levels.



High Low Moderate Non-Burnable Very High Very Low Water

Fig. 6.4 Layer Cam Visualization

In order, reduce computational effort while compromising features effectiveness, the DRFBs module incorporates a multi-branch approach built around the ResNet architecture. By which involves the hyper-parameter μ, the DRFBs module effectively reduces computational complexity, providing sub-branches to evaluate feature maps with decreased channel dimensions. Lightweight convolutional layers that operate with fewer parameters than conventional Res Net block represent the by-product of this technique. This allows for enhanced performance without getting the model deeper.

In contrast to traditional methods like the SE block, the DCFA module has an effective attention mechanism that keeps track of localized cross-channel encounters yet maintains dimension constant. The three-stage module proceeds through the following steps: first, it uses global average pooling to represent features globally; second, it uses one-dimensional convolution with a kernel size that is dynamically set; and third, it uses an appropriate activation function for normalization. The DCFA module optimizes channel attention prediction through improving the neural network more effective at learning using concentrating on local interactions.

Developing an accurate and cost-effective deep neural network for assessing fire risk becomes easier by incorporating all of these elements within the model architecture. Comprehensive risk assessments could be accomplished with less computational complexity by combining DRFBs and DCFA modules with a lightweight CNN architecture that uses multi-scale feature fusion. Improving the network model's performance and scalability, the proposed SSKD technique ensures successful training.

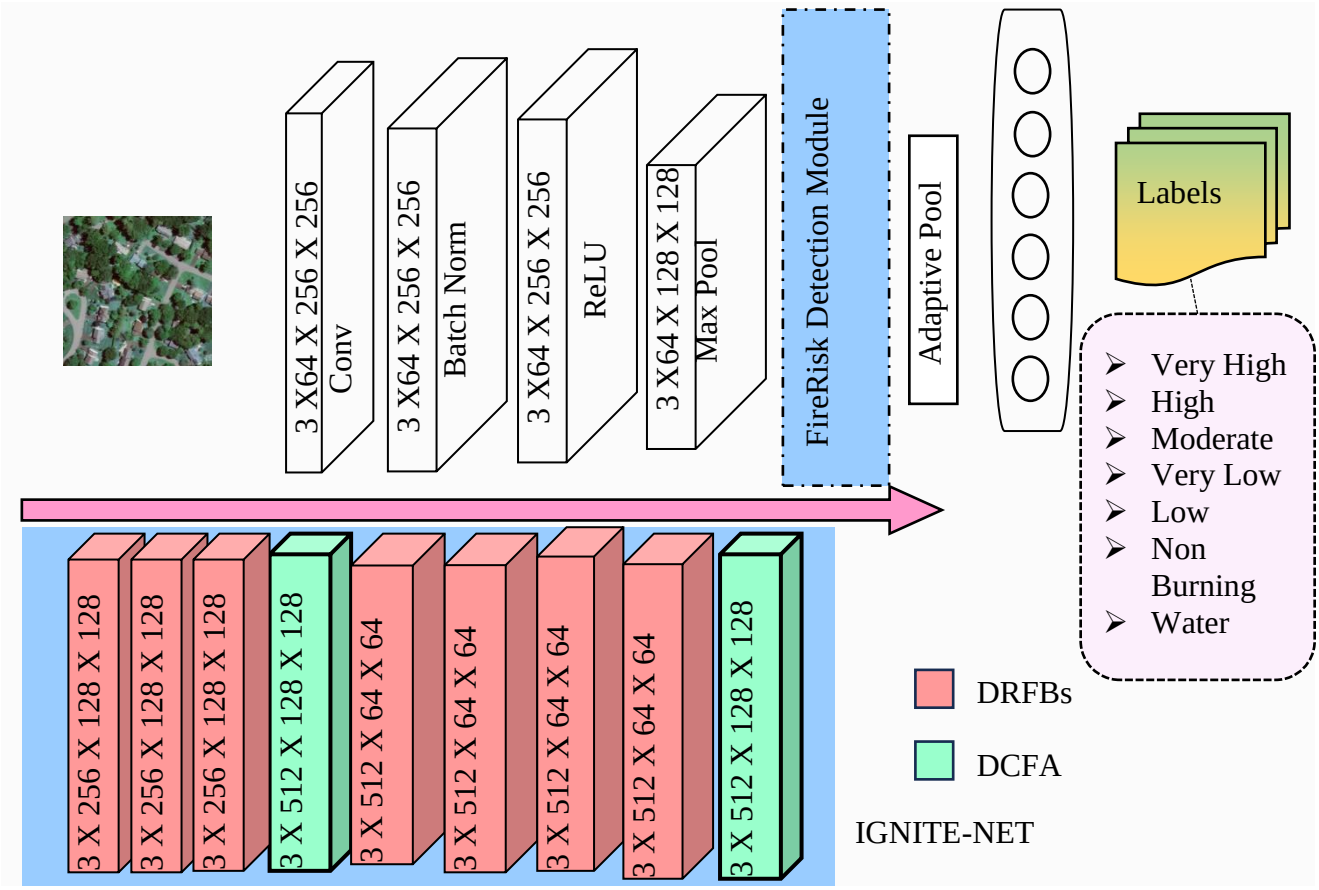


Fig. 6.5 Framework of the proposed Architecture where green block represents DCFA, and pink block represents DRFB.

The layer CAM representation in Fig. 6.4 visually highlights the model's concentration on specific regions of the input that contribute most to the classification decision. It also reveals the specific features within these regions, such as textures, shapes, or edges, that the model relies on to make its prediction. As depicted in

Table 6.1, the model framework uses sophisticated feature extraction algorithms and attention mechanisms. It also uses the FireRisk dataset, integrated with the Wildfire Hazard Potential (WHP) project. By combining the two data sets, we can evaluate fire risk prediction models, which helps us understand wildfire dynamics and develop better methods for controlling them.

6.2.3 Experimental Results and Discussion

The experimental methodology has been discussed in this section.

Dataset Selection and Pre-processing:

The experimental dataset serves a vital role in training and evaluating the effectiveness of the fire risk categorization model. A rigorously curated dataset of high-resolution photos depicting varied environmental situations prone to fire dangers has been selected for this investigation. The dataset is meticulously annotated, with each image labelled to represent one of seven discrete fire risk levels: 'high', 'low', 'moderate', 'non-burnable', 'very high', 'very low', and 'water'. The selection technique guarantees the equilibrium of classes and the inclusiveness of varied danger levels, encompassing fluctuations in lighting, weather, and environmental conditions to mirror real-life situations precisely.

Before model training, a sequence of pre-processing procedures is implemented to standardize and improve the dataset's appropriateness for training. The photos undergo conventional modifications, which involve scaling all images to a consistent resolution of 256x256 pixels and normalizing pixel intensities to a standardized level. Additionally, data augmentation techniques such as random horizontal and vertical flips enhance the training dataset, boosting the model's robustness and generalization capabilities.

Training Procedure

The proposed model is evaluated using F1-score, recall, accuracy, precision, and Matthew's Correlation Coefficient to assess the trained model. Table 1 clearly illustrates the performance metrics and the ranges of various indicators, showing the proposed model's robustness.

Accuracy: It depicts the model's prediction accuracy by comparing correctly categorized samples to the total samples in the dataset. An accurate representation may be imperfect, particularly in the case of class inequality.

Precision: This indicator shows the percentage of correct optimistic model projections. It is especially useful when false positives are costly. The percentage of optimistic projections among all true positive samples in the collection is known as recall, sensitivity, or actual positive rate. It evaluates the model's ability to capture all relevant instances of a particular class.

F1-score: The F1-score is the harmonic mean of precision and recall, providing a balanced measure of the model's performance. It is precious when there is an imbalance between the number of positive and negative instances in the dataset.

Matthews Correlation Coefficient (MCC): MCC is a correlation coefficient considering true and false positives and negatives. It ranges from -1 to 1, with 1 indicating perfect predictions, 0 indicating random predictions, and -1 indicating complete disagreement between predictions and ground truth.

Hardware and Software Environment

The experiments are conducted on a high-performance PC server equipped with an NVIDIA QUADRO RTX A5000 graphics card featuring a memory capacity of 24 GB and 3042 NVIDIA Cuda cores, acceleration to expedite model training and evaluation processes. The PyTorch deep learning framework is utilized for model implementation and experimentation, leveraging its extensive neural network development and training capabilities.

Result and Discussion

The evaluation metrics of various models, including HRNET, Resnext, Texture, DAVIT, SWIN_S, Max ViT, and our proposed model as a visual representation in Fig 6. , were meticulously analyzed to assess their performance in fire risk assessment. Each model has been

scrutinized based on test accuracy, Matthews's correlation coefficient (MCC), precision, recall, and F1 score.

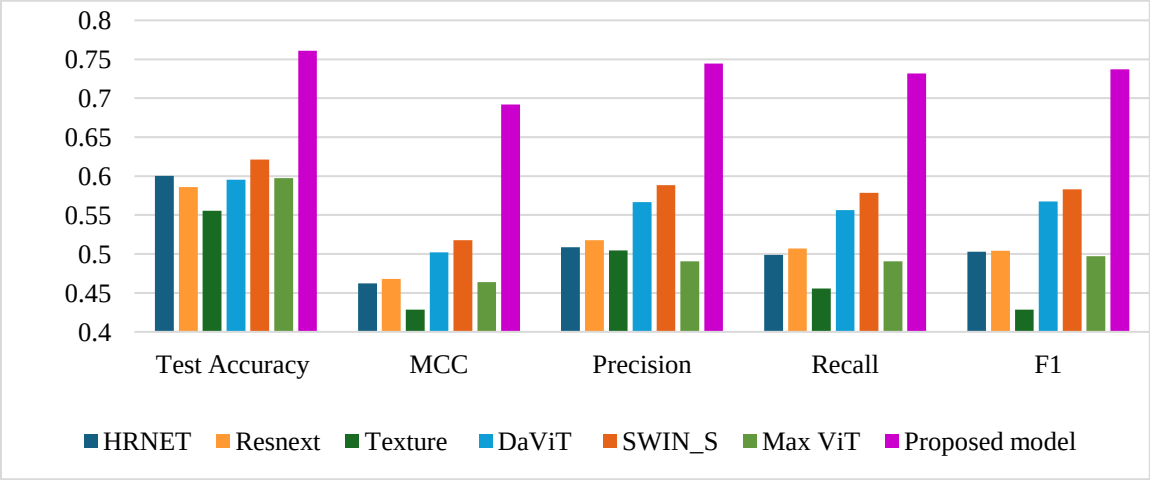


Fig. 6.6 Performance metrics comparison of the proposed model and latest SOTA model

The proposed model exhibited a test accuracy of 76.08%, outperforming other models such as HRNET (60.03%), Resnext (58.60%), Texture (55.56%), DAVIT (59.55%), SWIN_S (62.12%), and Max ViT (59.74%). This remarkable accuracy underscores the effectiveness of our model in accurately predicting fire risk levels.

Table 6.2 Performance Evaluation and Comparison with SOTA

Evaluation Metrics	HRNET	ReSNEXT	Texture	DAVIT	SWIN_S	Max ViT	**Proposed model**
Parameters(in Millions)	29	49	3	62	33	29	**31**
Test Accuracy	0.60	0.58	0.55	0.595	0.621	0.597	**0.760**
MCC	0.46	0.46	0.42	0.502	0.517	0.463	**0.692**
Precision	0.50	0.51	0.50	0.566	0.588	0.490	**0.744**
Recall	0.49	0.50	0.45	0.556	0.578	0.490	**0.731**
F1	0.50	0.50	0.42	0.567	0.583	0.497	**0.737**

Furthermore, the Matthews correlation coefficient (MCC) of our proposed model stood at 0.692, surpassing HRNET (0.462), Resnext (0.468), Texture (0.428), DAVIT (0.502), SWIN_S (0.518), and Max ViT (0.464). The high MCC value indicates a strong correlation between the

predicted and actual fire risk levels, highlighting the robustness of our model in capturing complex patterns and nuances in the dataset.

In terms of precision, recall, and F1 score, our proposed model demonstrated superior performance compared to other models. With scores of 0.744, 0.732, and 0.737 for accuracy, recall, and F1 respectively, our model achieved a balance between the detection of fire risk and the minimization of false positives.

Because of these scores, a proposed framework for fire risk assessment has been shown to be useful for environmental monitoring and natural disaster management. By utilizing state-of-the-art architecture and attention mechanisms, our model performs the fire risk more accurately than any SOTA model. The efficient performance of the proposed model contributes to the area of fire risk assessment, providing stakeholders with essential information that can be used to avoid wildfires and protect vulnerable ecosystems. Due to the robustness and dependability of our model, it is suitable for use in real-world scenarios for the purpose of developing proactive tactics to mitigate wildfires Overall, the results presented herein underscore the significant strides in fire risk assessment, propelled by cutting-edge research and innovative model architectures.

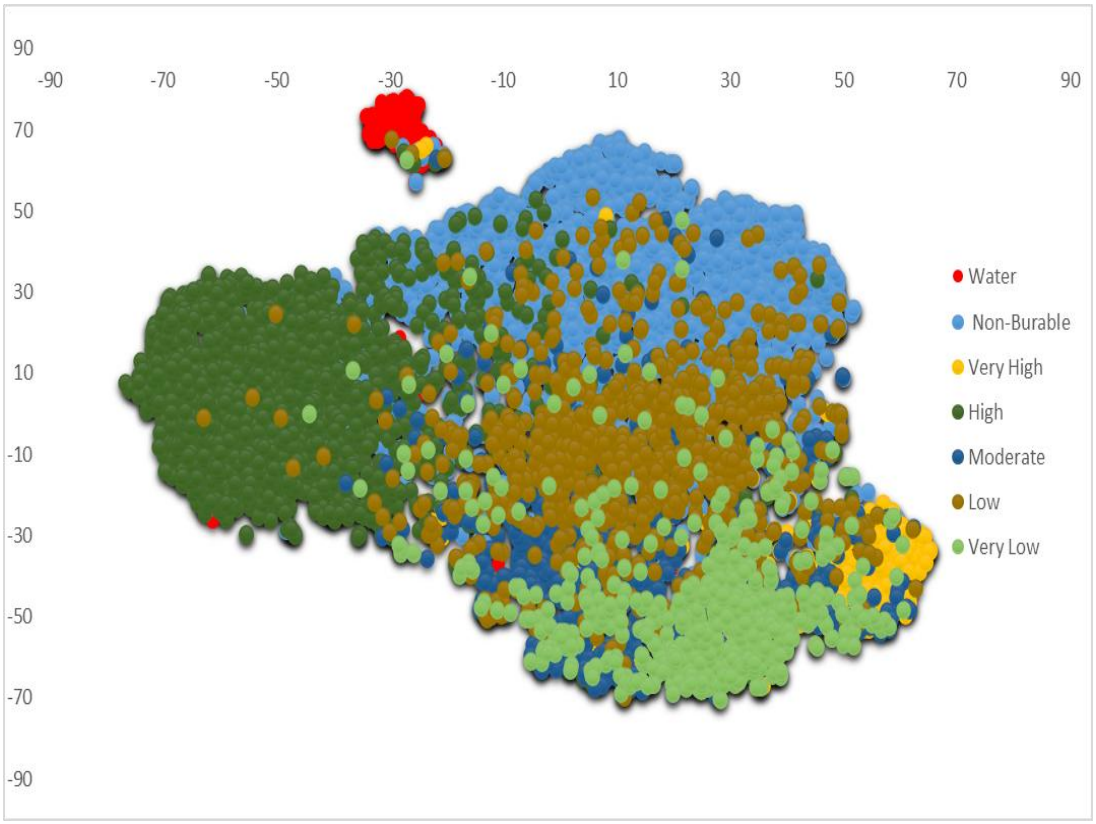


Fig. 6.7 t-SNE visualization depicting the distribution of FireRisk classes, highlighting the discriminative power of the proposed model in fire risk assessment

Analysis of T-SNE

The t-SNE portrayal in Fig gives a top-to-bottom comprehension of the spatial dissemination of FireRisk classes. This representation helps in understanding the discriminatory abilities of our proposed model. The significant clustering of fire risk classifications, illustrated in Fig. 7 above, demonstrates the model's ability to classify and distinguish between various fire hazard classes. This depiction provides a thorough overview of our model's performance in fire risk assessment. The proposed model's discriminative ability is better understood with the help of this visualization. This shows the unique clustering of fire risk categories, demonstrating the model's ability to capture and discriminate between various degrees of fire hazard. In addition to the numerical data, this visualization thoroughly reviews the ability of model performance in fire risk assessment.

6.2.4 Conclusion

Our research model is a novel framework with deep neural network architecture to predict fire-risk classes. The proposed approach identifies important categories of fire risk. The Wildfire Hazard Potential (WHP) initiative was significant in establishing and evaluating the

framework employing the FireRisk dataset. This model is designed to provide very accurate predictions of fire risk classes because it uses a lightweight convolutional neural network (CNN) architecture, a multi-scale feature extraction strategy, and a strong deep learning model called IGNITE-NET. We determined that our IGNITE-NET model was the most accurate compared to state-of-the-art models; therefore, we know it works. A high Matthews's correlation coefficient (MCC) demonstrated reliability in classifying fire risk classes in varied environmental scenarios, and our approach effectively balances recall, accuracy, and F1-score.

6.3 Significant Outcomes of this Chapter

The significant outcomes of this chapter are as follows:

- To enhance sea ice classification accuracy using SAR images through a novel hybrid deep learning model named "Arctic-Net". The proposed model integrates Convolutional Neural Networks (CNNs) and attention mechanisms, specifically the Adaptive Convolutional Encoder (ACE), Spatial Transposer Encoder (STE), and Hierarchical Transpose Attention (HTA), enabling efficient extraction of local and global features while maintaining computational efficiency.
- Conducted comprehensive performance evaluations using metrics such as accuracy, precision, recall, and F1-score. The Arctic-Net model achieved an accuracy of 0.93, precision of 0.91, and F1-score of 0.91, outperforming state-of-the-art models including DenseNet, ResNext, and Swin Transformer. These results demonstrate the model's superior classification capabilities and robustness.
- Demonstrated the practical applicability of Arctic-Net for operational sea ice monitoring, marine navigation, and climate research. The model's ability to accurately classify sea ice types with limited labeled data highlights its potential for real-time environmental monitoring and deployment in resource-constrained settings, contributing to advancements in climate studies and maritime safety.

The following research studies serve as the foundation for this chapter:

- ❖ Abhishek Verma, Virender Ranga, Dinesh Kumar Vishwakarma, "IGNITE-NET: Fire Risk Prediction using Dynamic Receptive Fields and Dynamic Channel Fusion Attention." Communicated in Advances in Space Research
- ❖ Indian patent published ,application number 202411073844 "System and Method for detecting Fire Prone Areas using an Unmanned aerial Vehicle"

This chapter introduces IGNITE-NET, an innovative deep learning framework designed to predict fire risk levels by leveraging dynamic receptive field blocks (DRFBs) and dynamic channel fusion attention (DCFA). The model significantly reduces computational complexity while achieving superior predictive accuracy, demonstrating its potential for real-time fire risk assessment.

The next chapter explores the performance versus computational complexity trade-off in fire risk detection, focusing on the Swin Transformer architecture and evaluating models with RGB and edge-based inputs. This study aims to provide insights into the balance between model performance and computational cost in cross-domain fire risk detection.

Chapter 7 : OPTIMIZING FIRE RISK DETECTION: BALANCING MODEL PERFORMANCE VS COMPUTATIONAL COST

7.1 Scope of this Chapter

This chapter delves into the intricate balance between performance and computational complexity in cross-domain fire risk detection, using advanced machine learning models, particularly focusing on the Swin Transformer architecture. The study emphasizes evaluating models with varying input types—RGB-based and edge-based images—derived from the FireRisk dataset, a comprehensive collection of remote sensing imagery specifically designed for fire risk assessment.The chapter is structured to provide a thorough exploration of the methodologies, from data pre-processing to model selection and performance evaluation. It begins by detailing the pre-processing techniques applied to the RGB and edge-based images, highlighting how these inputs influence model training and accuracy. The selection of vision models, including Swin Transformer, HRNet, ResNeXt, Max ViT, and a Texture-specific ResNet50, is discussed in the context of their architectural advantages and limitations concerning fire risk detection. A significant portion of the chapter is dedicated to the experimental setup, encompassing hardware specifications, training parameters, and data augmentation strategies. The performance of each model is meticulously analyzed through various metrics such as accuracy, Matthews Correlation Coefficient (MCC), precision, recall, and F1-score, providing a comprehensive understanding of each model's capabilities and limitations.The chapter also includes a comparative analysis with state-of-the-art models to underscore the advancements achieved through the proposed methodologies.

Furthermore, it explores the generalization capabilities of these models across different datasets, emphasizing the importance of robust model training for effective real-world application.

In conclusion, this chapter not only presents a detailed examination of model performance and computational trade-offs in fire risk detection but also sets the stage for future research

directions, including the integration of temporal data, ensemble learning, and the development of real-time detection systems.

7.2 Investigating the Performance vs Computational Complexity Tradeoff in Cross-Domain Fire Risk Detection

7.2.1 Abstract

Fire risk detection is critical for timely interventions and effective management strategies in mitigating wildfire impacts. This study examines the efficacy of many advanced models, emphasizing the Swin Transformer architecture for efficient fire detection. We assessed RGB input evaluations, highlighting the Swin_S model, which attained a test accuracy of 62.1%, and the Swin_T model at 61%. Comparative analysis with current models demonstrated that Swin_T_Edge surpassed its competitors, achieving the maximum accuracy of 66% and an F1 score of 0.587, confirming its efficacy in classification tasks while maintaining a balance in model complexity. Cross-dataset tests further illustrated the models' durability across various fire conditions, underscoring the necessity for solid generalization capabilities in real-world applications. Statistical evaluations utilizing t-tests confirmed the substantial performance enhancements of the suggested models. The findings highlight the Swin_T_Edge model's promise as a premier option for fire risk detection systems, recommending future improvements via ensemble learning and the incorporation of temporal data

7.2.2 Proposed Methodology

The proposed methodology is discussed in this section.

Dataset Description

The FireRisk dataset , is a carefully compiled collection of remote-sensing photos for fire risk evaluation. It is a crucial resource for developing and accessing fire risk prediction models. The FireRisk dataset was not generated for this study; instead, it is employed to determine the efficacy of different prediction algorithms.

The data for this research originates from the Wildfire Hazard Potential (WHP) project, established by the U.S. Department of Agriculture, recognized for its comprehensive evaluations of fire risk and wildfire intensity throughout the United States. The WHP project integrates geostatistical data sources, such as FSim for assessing wildfire susceptibility and severity, LAND-FIRE for fuel and vegetation information, and the Fire Program Analysis

(FPA) for historical fire occurrence records. The 2020 version of the WHP raster dataset offers detailed fire risk evaluations classified into seven specific categories.

The raster dataset is provided in a geodatabase format (.gdb) and divides the nation into grids, each measuring 270 meters per side, along with associated fire risk assessments for each grid cell. The images in the FireRisk dataset are obtained from the National Agriculture Imagery Program (NAIP), which utilizes airborne platforms to capture high-resolution ortho-rectified imagery with a spatial resolution of no less than 1 meter, exceeding the quality generally attained through satellite-based remote sensing. Strict quality criteria regulate the images, necessitating a sun elevation of no less than 30 degrees and a maximum cloud cover of 10% for each quarter of the image segments. Images are gathered during the growing season to reduce the occurrence of snow and flooding. The collection comprises 91,872 remote-sensing photos of fire risk assessments derived from the WHP dataset. Of them, 70,331 photos are designated for training, whereas 21,541 images are allotted for validation. Every image is subjected to a uniform cropping procedure, yielding dimensions of 270 × 270 pixels. It is classified into seven discrete fire risk categories, enabling a comprehensive analysis of fire risk levels across diverse geographical regions.

RGB-Based Input

The initial data preparation scenario involves a photograph dataset based on the RGB color model. The images are utilized in an unaltered color format without image filtering or alteration techniques. The preparation procedures for this subset of the dataset encompass the subsequent steps:

Image Loading: The dataset is retrieved from the designated directory and consists of images in RGB format.

Label Assignment: Every image is labeled according to the class in the image's file name. The designation is employed in tasks related to supervised learning.

Image Transformation: The images undergo a transformation process to achieve a standardized size, typically 256x256 pixels, to maintain consistency throughout the training phase.

Data Augmentation: Data augmentation strategies enhance the model's generalization. These transformations encompass random horizontal flips and random vertical flips.

Edge Based Input

In the second data preparation scenario, the dataset comprises images based on edges. Edge-based photos are produced by applying an edge-detection filter, such as the "FIND_EDGES" filter, on the initial RGB images. The preparation procedures for this particular subset of the dataset are outlined as follows:

Image Loading: The dataset, which consists of images that have undergone edge filtering, is imported from the designated directory. *Label Assignment*: Just like RGB-based images, edge-based images are assigned a label depending on the class indicated in the image's file name.

Edge Filtering: The edge-filtered images are obtained by applying the "FIND_EDGES" filter on the original RGB images. This procedure improves the perceptibility of boundaries and outlines inside the pictures.

Image Transformation: Like RGB images, the edge-based images undergo resizing to achieve a standardized size, guaranteeing uniformity throughout the model training process.

*Data Augmentation**:*** Data augmentation strategies are employed for edge-based images to bolster the model's resilience. The strategies encompass the utilization of random horizontal flips and random vertical flips.

The Preprocessing dataset is divided into RGB-based and edge-based images; the data preparation pipeline can effectively accommodate diverse modeling techniques. This allows for considering either raw color information or the prioritization of edge characteristics, depending on the specific modeling requirements. The subdivision enables the creation of specialized models that can efficiently acquire knowledge from several image categories to tackle specific parts of your study, such as evaluating fire hazards

Pre-processing of input types

The dataset preparation encompasses two input types: RGB pictures and edge images. This structured data preparation pipeline supports several modelling strategies, using raw color data or focusing on edge features based on individual analytical needs. Utilizing these two input sources in succession enables the development of models that effectively learn from each image category, improving our capacity to evaluate fire danger levels across various geographical

areas. This method enhances the models' flexibility and guarantees a thorough assessment of fire dangers.

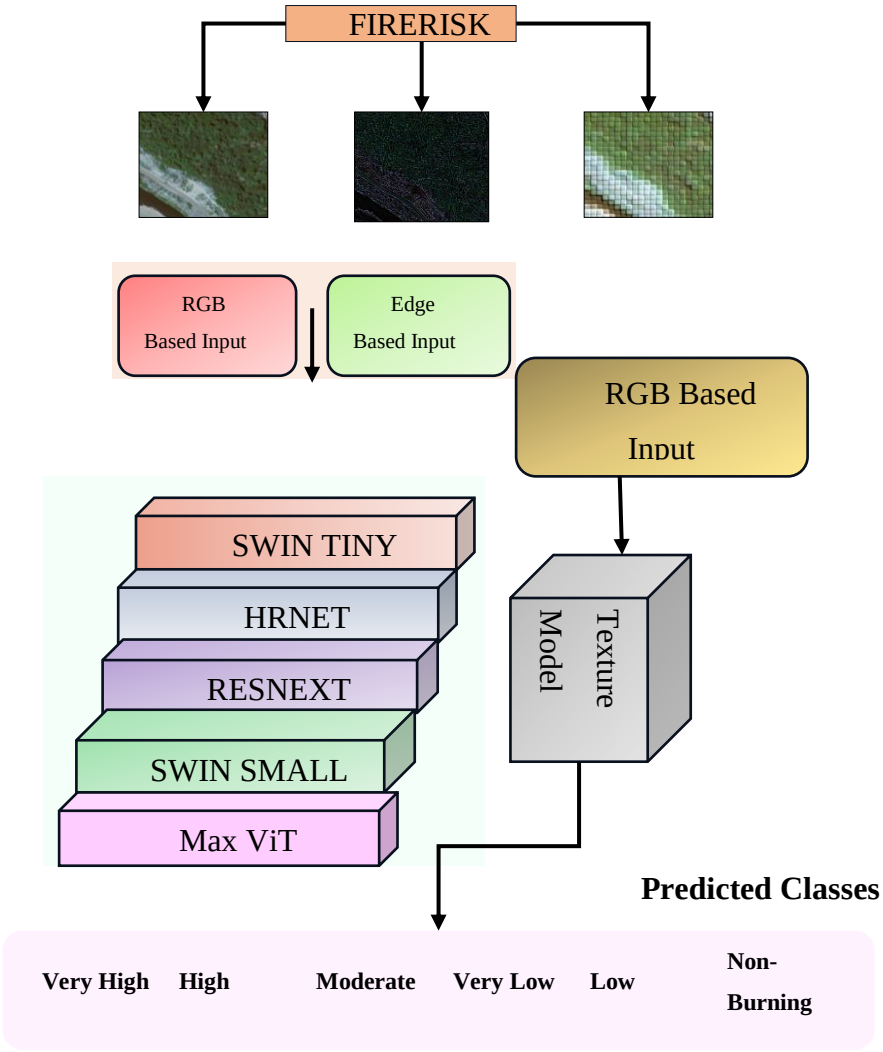


Fig. 7.1 Proposed Framework for Fire Risk Assessment, illustrating the processing of RGB and edge-based inputs for classifying fire risk levels.

Model Selection

The methodology utilized in the study for detecting fire danger across different domains involves utilizing several vision models to analyze remote-sensing images. The models have been chosen based on their appropriateness for Image classification tasks, namely those that utilize edge information, RGB, and exclusive texture analysis. The model selection process encompasses.

Proposed framework

This Section explains the proposed framework flow and model description employed for further analysis, as can also be observed in Fig. 7.1.

SWIN Transformer

The Swin Transformer (Swin_T_Edge)[76] is a significant breakthrough in visual modeling. This model presents advancements in vision transformers by including innovative architectural elements, like multi-headed self-attention with a window-based mechanism and shifting windows. These advancements augment Swin's ability to capture distant relationships in visual information efficiently. The training process employed by Swin closely aligns with that of the Data Efficient Image Transformer (DeiT), with a particular emphasis on efficiency. The process starts with pre-training the ImageNet-21k dataset and then fine-tuning the ImageNet-1k dataset [84]. Swin's performance is outstanding, achieving a new state-of-the-art benchmark on the Tiny ImageNet dataset. Notably, it has achieved a validation accuracy of 91.35%, exceeding the previous leading model by 0.33%. Swin's adeptness in managing the Tiny ImageNet dataset, in combination with its distinct window-based attention mechanism and the accessibility of its source code for additional investigation, establishes it as an essential selection for image classification tasks and a pivotal point of citation for researchers in this domain. The study also alludes to additional transformer versions, such as Swin and MaxViT, presenting intriguing prospects for further progressions in vision transformer models.

HRNET

The HRNet architecture, first developed for human posture estimation, has demonstrated its versatility and applicability in several computer vision tasks. HRNet is proficient in maintaining high-resolution representations, an essential prerequisite for functions that need intricate spatial information, including semantic segmentation, facial landmark identification, and object detection.

HRNet does this by maintaining several parallel convolutions that cover a range of resolutions, from high to low. Additionally, it consistently integrates multi-scale information across these parallel streams through fusion techniques comparable to model sizes and computational efficiency[85].

RESNEXT

The ResNeXt architecture is a sophisticated convolutional neural network (CNN) structure that builds upon the foundational ideas of Residual Networks (ResNets). The text presents the fundamental notion of "cardinality," which serves as a metric for quantifying the quantity of concurrent pathways inside a given network. These parallel routes, commonly known as "cardinalities," might be seen as a collective effort of numerous specialists working together to

address a problem. By integrating this notion, ResNeXt enables the network to get a broad spectrum of varied and comprehensive feature representations.

The cardinality-driven design of ResNeXt proves to be particularly helpful in the context of your fire risk dataset. The system has exceptional proficiency in collecting delicate and nuanced characteristics of utmost importance for applications such as image categorization about fire hazards. ResNeXt's remarkable feature extraction skills enable precise and accurate operation of your model, whether recognizing fire dangers in images or finding subtle patterns indicative of possible risk factors[70].

Multi-Axis ViT (Max Vit)

The Max ViT architecture represents a novel implementation of the vision transformer (ViT) model, characterized by its efficiency and scalability. The proposed approach incorporates a multi-axis attention mechanism, enabling the model to capture global-local spatial interactions across various input resolutions while maintaining linear computational cost. Moreover, the Max ViT model integrates convolutional layers into its design to enhance efficiency. The Max ViT model has exceptional performance on many image classification benchmarks, including ImageNet-1K, ImageNet-21K, and CIFAR-100, establishing itself as the current leader in the field. Additionally, it has robust scalability when used for datasets of considerable size and high-resolution images[86].

In conclusion, Max ViT has considerable strength and adaptability as a ViT model, holding promise as a future frontrunner in many computer vision applications. The layer CAM presented in Fig. 7.2 highlights the regions where the RESNEXT model focuses its attention.

TEXURE Model

The TEXTURE ResNet50 model is a variation of the ResNet50 architecture specifically designed to cater to the requirements of computer vision jobs. The term "TEXTURE" in the model's nomenclature alludes to its distinct emphasis on analyzing texture inside images. The architecture of this model is optimized explicitly for cases in which the proper analysis of images heavily relies on texture-based information.

TEXTURE ResNet50 inherits the fundamental structure of the ResNet50 architecture at its core. The ResNet50 architecture is a convolutional neural network (CNN) comprising 50 layers. This depth gives it a greater capacity for learning complex hierarchical features than earlier models. The fundamental structure of the architecture consists of a sequence of

convolutional layers, residual blocks, and fully linked layers. The combined elements of these components contribute to the model's capacity to effectively capture intricate patterns seen in images.

One distinguishing characteristic of TEXTURE ResNet50 is its notable focus on examining texture. The system is designed to effectively identify and analyze texture patterns present in images. Examining texture is pivotal in image content analysis, particularly in material identification, surface examination, and specific medical imaging assignments. The specialty of TEXTURE ResNet50 allows it to perform exceptionally well in jobs requiring a high level of emphasis on comprehending intricate aspects of texture[87].

Comparison of Methodologies

This section provides a succinct comparison of the approaches examined, emphasizing their advantages and drawbacks. Each strategy offers distinct benefits for fire risk detection while also posing obstacles. The table below delineates the principal advantages and disadvantages of each model.

Table 7.1 delineates the various strengths and limitations of the models in this comparison. The Swin Transformer is distinguished by its exceptional accuracy, but HRNet is superior at jobs necessitating spatial precision. ResNeXt's cardinality-centric methodology is proficient in intricate feature extraction, whereas Max ViT provides a scalable resolution. The Texture model is optimized for texture analysis, rendering it highly appropriate when texture is paramount. Each methodology offers distinct advantages for fire risk detection. Their choice is contingent upon the particular demands of the work, including the necessity for texture analysis, computing efficiency, or high-resolution representation.

Table 7.1 This table concisely compares the models, focusing on their key strengths and limitations relevant to fire risk detection tasks.

Model	Strengths	Limitations
Swin Transformer	Efficient multi-scale attention State-of-the-art performance on classification tasks	High computational complexity
HRNet	Preserves high-resolution details Excellent for tasks requiring spatial precision	High computational demand due to multi-resolution processing
ResNeXt	Robust feature extraction via cardinality	Increased complexity leads to longer training times

	Effective for fine-grained pattern detection	
Max ViT	Efficient and scalable	Limited real-world validation beyond benchmark datasets
	Strong performance on large-scale benchmarks	
Texture	Optimized for texture analysis	Limited focus on broader contextual features
	Based on proven ResNet50 architecture	

7.2.3 Experimental Results and discussion

This section presents the experimental setup employed in the study, followed by a comprehensive analysis and comparison of the performance of various vision models for cross-domain fire risk detection. The results are evaluated based on key performance metrics, highlighting the effectiveness and efficiency of each model in detecting fire risks across diverse datasets.

Hardware Configuration

The experiments used a high-performance workstation with two NVIDIA A5000 graphics cards, an Intel Xeon processor, and 128 GB of RAM. The hardware configuration was chosen to optimize the model training and evaluation process.

Training Details

Data Parallelism: Data parallelism was employed to distribute the training workload efficiently across the two NVIDIA A5000 graphics cards. This approach optimized training times.

*Batch Size***:** A batch size of 64 balanced training efficiency and memory utilization.

*Number of Epochs***:** 50 epochs were used for training to guarantee model stability and convergence. After conducting multiple rounds of experiments, it was determined that training for 50 epochs was sufficient to achieve optimal model performance and convergence.

*Learning Rate Schedule***:** The learning rate was adjusted using the cosine annealing technique, facilitating efficient model convergence.

Optimizer: The most recent iteration of the AdamW optimizer, renowned for its potency in deep neural network training, is employed.

Data Preprocessing

The dataset was partitioned into two distinct groups, namely RGB-based images and edge-based images. The preparation operations for RGB-based images encompassed many stages, namely importing the images, assigning appropriate labels, scaling them to a specified dimension of 256x256 pixels, and using data augmentation techniques such as random horizontal and vertical flips. To improve the visibility of boundaries in edge-based images, a filter called "FIND_EDGES" was utilized for edge detection. Subsequently, identical pre-processing procedures were employed for RGB-based images.

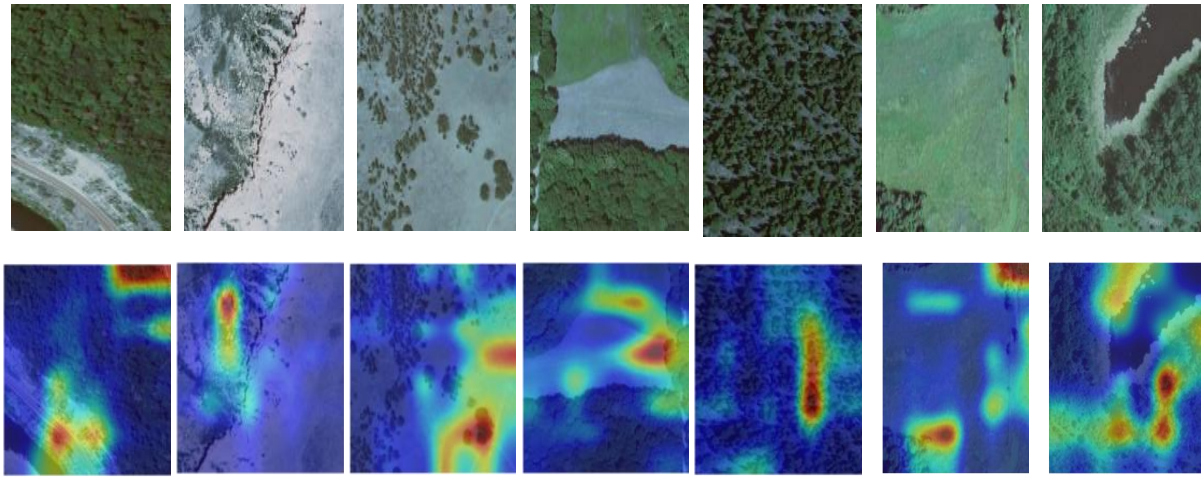


Fig. 7.2 Original image and Layer CAM Visualization in ResNeXt Original image (top) alongside Layer CAM visualization (bottom). The Layer CAM highlights regions of interest in the original image, providing insights into the model's focus areas during classification, categorized into five classes: High, Low, Moderate, Non-Burnable, Very High, Very Low, and Water.

Model Performance and Complexity Analysis

This section thoroughly examines model performance and computational complexity to provide significant insights into their efficacy in cross-domain fire risk detection while considering computing requirements.

As presented in Fig. 7.3, the findings provide a comprehensive overview of key performance indicators for each model. These indicators encompass CPU Time, GPU Time, Multiply-Accumulate (MAC) operations, parameter count, and accuracy. This comprehensive research enables well-informed conclusions on the trade-offs between a model's performance and computational complexity.

The research reveals that the SWIN_T_Edge model attains the highest level of accuracy, measuring 0.66. It is closely trailed by the SWIN_S model, which reaches an accuracy of 0.62.

These models exhibit a noteworthy level of accuracy while simultaneously keeping the number of parameters manageable. In contrast, the Texture Model demonstrates a moderate level of accuracy, precisely 0.55, while exhibiting a notably low level of computing complexity. This underscores the promise of texture-based models within this field.

Nevertheless, it is crucial to consider the computing demands associated with these models. The Max ViT model, which consists of 64.021 million parameters, exhibits significant computational complexity, specifically about "CPU time" and "GPU time". It can be observed in Table 7.2. Achieving an optimal trade-off between performance and complexity is paramount in scenarios with limited resources.

In addition to the tabular data , A bubble chart in Fig. 7.3 is utilized to visually illustrate the relationships among model accuracy, the number of parameters, and the Multiply-Accumulate Operations (MAC). In this graph, the y-axis represents accuracy, while the x-axis represents the number of parameters. The size of each bubble corresponds to the MAC, providing a clear representation of how these metrics interact.

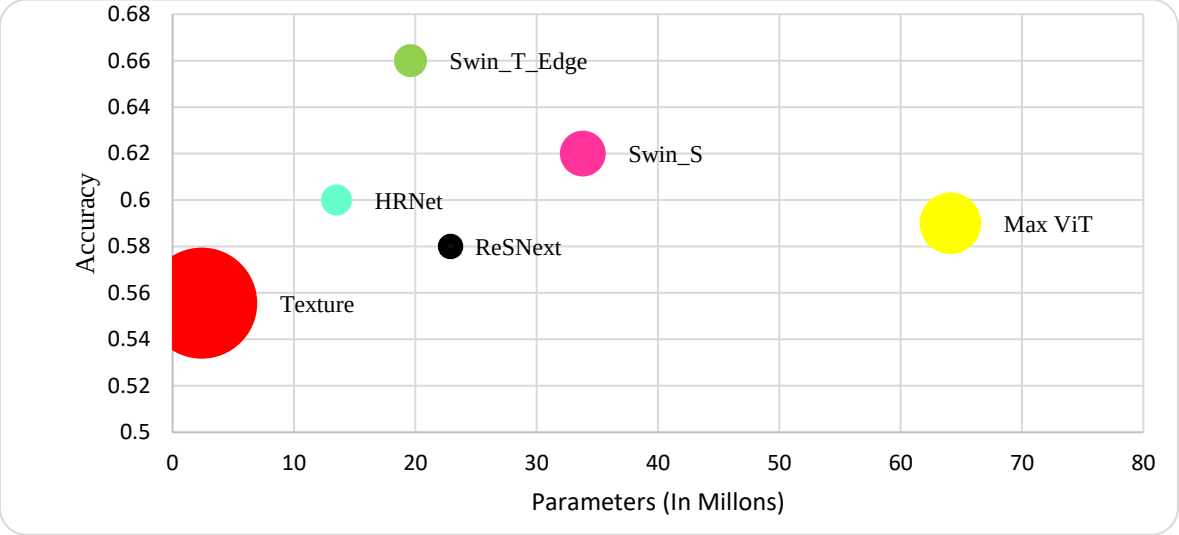


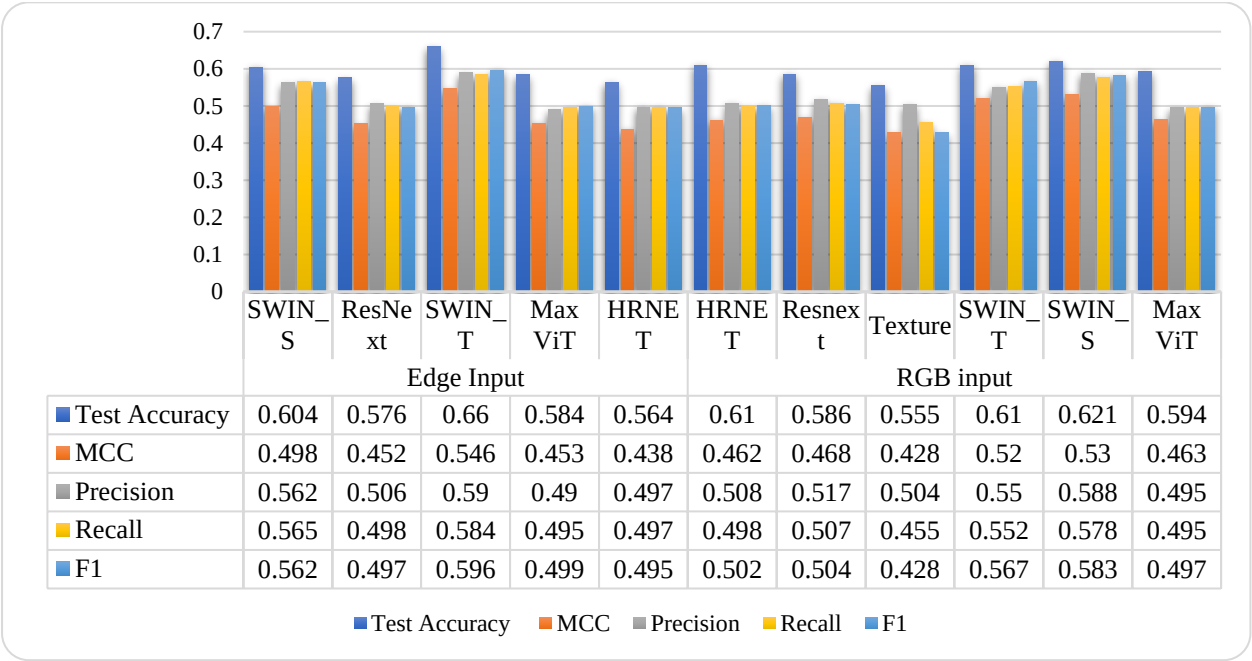
Fig. 7.3 Model Performance Trade-off. This figure illustrates the relationship between model complexity and accuracy, where the size of each bubble corresponds to the number of parameters (in millions) for each model, effectively demonstrating the trade-off between computational demands and performance metrics.

Result and Discussion

This part thoroughly examines model performance, specifically emphasizing diverse evaluation measures for the models across distinct input situations, namely Edge Input and RGB Input.

The evaluated models include SWIN_T_EDGE, ResNext, SWIN_S, Max ViT, HRNET, and Texture under RGB Input circumstances. Comprehensive performance analysis using essential measures such as Test Accuracy, MCC (Matthews Correlation Coefficient), Precision, Recall, and F1-score.

Fig. 7.4 presents a comprehensive summary of the performance metrics for each model utilized in the study. The findings reveal considerable heterogeneity in CPU and GPU processing durations, with the quantity of Multiply-Accumulate Operations (MAC) and parameters. The Swin_T_Edge model attained the maximum accuracy of 0.66 while preserving a comparatively low computational expense, indicating it may be the most efficient option for fire risk assessment applications among the assessed models.



	SWIN_S	ResNext	SWIN_T	Max ViT	HRNET	HRNET	Resnext	Texture	SWIN_T	SWIN_S	Max ViT
	Edge Input					RGB input					
Test Accuracy	0.604	0.576	0.66	0.584	0.564	0.61	0.586	0.555	0.61	0.621	0.594
MCC	0.498	0.452	0.546	0.453	0.438	0.462	0.468	0.428	0.52	0.53	0.463
Precision	0.562	0.506	0.59	0.49	0.497	0.508	0.517	0.504	0.55	0.588	0.495
Recall	0.565	0.498	0.584	0.495	0.497	0.498	0.507	0.455	0.552	0.578	0.495
F1	0.562	0.497	0.596	0.499	0.495	0.502	0.504	0.428	0.567	0.583	0.497

Fig. 7.4 This figure visually represents the evaluation metrics, including Test Accuracy, Matthews Correlation Coefficient (MCC), Precision, Recall, and F1 score for various models (SWIN_T, ResNext, SWIN_S, Max ViT, HRNET) using edge and RGB inputs. The chart illustrates the performance differences across models, clearly comparing how each model performs with different input types.

Table 7.2 Performance Metrics of Models table presents the CPU and GPU time per epoch, Multiply-Accumulate Operations (MAC), number of parameters (in millions), and accuracy (ACC) for each model, offering a concise comparison of their computational efficiency and performance.

Models	CPU Time per epoch	GPU Time per epoch	MAC	PARAM's (in Millions)	ACC
ResNEXT[79]	0.31	0.7	1.399	22.994	0.58
Swin_S[80]	0.45	0.6	240.73	33.818	0.62
HRNET[88]	0.22	0.68	110.623	13.562	0.55
Texture[87]	21.48	0.8	1412.74	2.389	0.59
Max ViT[89]	1	0.19	434.665	64.021	0.59
Swin_T_Edge	0.29	0.3	124.464	19.621	0.66

Edge Input Evaluation

The assessment under edge input conditions indicates that SWIN_T consistently surpassed other models, attaining the highest Test Accuracy of 66%. This signifies an enhanced capacity to categorize edge-based inputs accurately. SWIN_S achieved an accuracy of 60.4%, indicating a competitive performance. Max ViT and ResNext attained modest accuracies of 58.4% and 57.6%, respectively, demonstrating their efficacy, although lacking the performance of the top models. Regarding the Matthews Correlation Coefficient (MCC), which assesses the effectiveness of classifications, SWIN_T scored 0.546, highlighting its dependability in extreme situations, whereas SWIN_S recorded 0.498. In terms of Precision, indicating the capacity to identify positive instances accurately, SWIN_T (0.59) and SWIN_S (0.562) exhibited the highest accuracy, but Max ViT (0.49) and HRNET (0.497) showed marginally lesser accuracy in recognizing genuine positives. The Recall was the highest, indicating the model's ability to identify all positive instances.

SWIN_T (0.584) and SWIN_S (0.565) demonstrated their efficacy in detecting positive cases, while Max ViT and HRNET followed closely at 0.495 and 0.497, respectively. In integrating accuracy and recall inside the F1-score, SWIN_T attained the optimal equilibrium (0.596), succeeded by SWIN_S (0.562), establishing these models as the foremost contenders under edge input conditions. A detailed illustration is in Fig. 7.4.

RGB Input Evaluation

In the RGB input assessment, SWIN_S surpassed other models with a Test Accuracy of 62.1%, underscoring its exceptional capability to process RGB inputs. SWIN_T was closely behind at 61%, and Max ViT attained 59.4%. This demonstrates the capacity of SWIN_S and SWIN_T to generalize effectively over RGB input data. The MCC scores corroborated these

findings, with SWIN_S achieving the maximum score of 0.53 and SWIN_T at 0.52, signifying a robust association between anticipated and actual classifications. Max ViT exhibited a reduced MCC of 0.463, indicating marginally less consistent performance. The maximum precision was achieved by SWIN_T at 0.588, indicating its superior capability in recognizing true positives, but SWIN_S also demonstrated commendable performance with a precision of 0.55. Max ViT and HRNET achieved accuracy scores of 0.495 and 0.504, respectively, signifying an elevated false positive rate relative to the leading models. Recall scores were highest for SWIN_S (0.578) and SWIN_T (0.552), indicating their efficacy in identifying positive instances, while Max ViT and HRNET followed at 0.495 and 0.498, respectively. The F1-score, which reconciles precision and recall, was highest for SWIN_S (0.583), followed by SWIN_T (0.567), underscoring their exceptional performance with RGB inputs.

Comparison with State of the Arts

Table 7.3 presents a comparison of various state-of-the-art models based on accuracy (Acc), F1 score (F1), and parameter count (in millions). Swin_T_Edge excels with the highest accuracy of 0.66 and an F1 score of 0.587, demonstrating its superior performance in classification tasks. In contrast, Dense-net, EfficientNet-B0, and MobileNetV3-Large exhibit lower accuracy and F1 scores, reinforcing the effective balance that Swin_T_Edge achieves between performance and model complexity. This highlights Swin_T_Edge as the most effective choice for applications requiring high accuracy and efficiency.

Table 7.3: Comparison with State-of-the-Art Models -Accuracy (Acc), F1 score (F1), and parameters (millions) for different state-of-the-art models, showing the balance between performance and model size.

Model	**Acc**	**F1**	**Parameters**
Dense-net[77]	0.55	0.49	~24M
EfficientNet-B0[90]	0.53	0.47	~5.3M
MobileNetV3-Large[91]	0.58	0.51	~5.4M
DenseNet-121[92]	0.59	0.52	~8M
ConvNeXt-Tiny[93]	0.6	0.53	~28M
ViT[94]	0.612	0.501	~86M
DINO[95]	0.628	0.526	~87M
MAE[96]	0.633	0.549	~87M
Swin_T _Edge	0.66	0.587	~28M

Generalization and Cross-Dataset Evaluation

This section assesses the generalization capabilities of the proposed models over diverse datasets to gauge their robustness and adaptability in various real-world fire scenarios. We conducted a cross-dataset evaluation, wherein models trained on one dataset are assessed on the other, as outlined in Table 7.4. This evaluation is essential for determining the generalization capability of models when confronted with unfamiliar data exhibiting varying features, which is vital for fire detection systems that must function in multiple contexts and situations.

Table 7.4 illustrates that models, including ResNEXT, Swin_T_Edge, Swin_S, HRNET, and Max ViT, were trained and evaluated on both the FD and YAR datasets[97], [98]. The outcomes are delineated in terms of accuracy (ACC), precision (P), recall (R), and F1-score (F1). The models demonstrate differing levels of performance based on the training and testing combinations employed. Typically, models evaluated on the same dataset exhibit superior accuracy and enhanced performance measures relative to cross-dataset evaluations, underscoring the significance of dataset variety in cultivating strong models.

For example, ResNEXT attained an accuracy of 0.87 when trained and evaluated on FD but a diminished accuracy of 0.79 when assessed on YAR, highlighting the difficulties in cross-dataset generalization. Comparable tendencies are noted with several models, including Swin_T_Edge and HRNET. Max ViT has consistently superior performance across several datasets, rendering it an appropriate choice for fire detection in diverse situations. This assessment offers essential insights into the generalization capacities of the suggested models, emphasizing the need for cross-dataset validation in fire risk detection systems.

Table 7.4 Generalization Study on the datasets where X represents FD and Y represents YAR. This table demonstrates the cross-dataset evaluation of models trained on FD[97] and YAR[98] datasets, highlighting their performance in terms of accuracy (ACC), precision (P), recall (R), and F1-score (F1) when tested across the two datasets.

MODELS	TRAIN	TEST	ACC	P	R	F1
ResNEXT[79]	X	X	0.87	0.84	0.89	0.88
	Y	Y	0.89	0.85	0.9	0.89
	X	Y	0.79	0.8	0.81	0.82
	Y	X	0.77	0.78	0.79	0.81
Swin_S[80]	X	X	0.91	0.9	0.89	0.89
	Y	Y	0.92	0.87	0.9	0.88

	X	Y	0.85	0.83	0.85	0.84
	Y	X	0.84	0.8	0.79	0.8
HRNET[88]	X	X	0.88	0.86	0.89	0.87
	Y	Y	0.87	0.84	0.86	0.85
	X	Y	0.85	0.82	0.84	0.83
	Y	X	0.84	0.83	0.85	0.84
Max ViT[89]	X	X	0.92	0.91	0.92	0.91
	Y	Y	0.93	0.89	0.91	0.9
	X	Y	0.88	0.84	0.89	0.85
	Y	X	0.88	0.85	0.86	0.84
Swin_T_Edge	X	X	0.89	0.88	0.87	0.87
	Y	Y	0.9	0.86	0.88	0.87
	X	Y	0.86	0.81	0.84	0.85
	Y	X	0.81	0.79	0.79	0.78

Evaluating Statistical Superiority: T-Test Analysis of Model Performance

The t-test is a robust statistical technique employed to evaluate the differences in means between two samples, enabling researchers to ascertain if observed variances are statistically significant. This study used a thorough t-test to assess the efficacy of our suggested technique compared to several leading models. The findings display the normalized t-statistics and p-values for pairwise comparisons. The Swin_T_Edge model regularly exhibits higher accuracy than others, with p-values reflecting substantial statistical significance ($p < 0.05$) in all comparisons. This indicates that the performance improvements realized by our method are improbable to be coincidental, hence substantiating its efficacy in fire detection tasks. The null hypothesis (H_0) states that no substantial difference exists between the means of the models being compared, whereas the alternative hypothesis (H_a) asserts that such differences exist. Employing a significance threshold of 0.05, our investigation underscores the resilience of the Swin_T_Edge model, positioning it as a premier candidate in fire detection techniques.

7.2.4 Conclusion

In summary, the results of this study offer significant contributions to our understanding of the complicated interplay between the performance and complexity of vision models when applied to cross-domain fire risk detection. This paper thoroughly assesses many advanced fire risk detection models, emphasizing the effectiveness of the Swin_T_Edge architecture. The

experimental findings indicated that Swin_T_Edge attained the maximum accuracy (66%) and an F1 score (0.587), surpassing traditional models while preserving a favorable equilibrium between performance and model complexity. Examining RGB input data validated the enhanced generalization skills of SWIN_S and SWIN_T, attaining substantial metrics affirming their efficacy in fire detection tasks. Moreover, the cross-dataset evaluation emphasizes the need for rigorous model training across diverse datasets to improve flexibility in practical applications. The statistical analysis, bolstered by t-test assessments, confirms the superiority of the presented models, particularly the Swin_T_Edge, highlighting its potential as a premier solution in fire risk assessment.

7.3 Significant Outcomes of this Chapter

The significant outcomes of this chapter are as follows:

- This study optimized cross-domain fire risk detection using advanced vision models, with the Swin_T_Edge model achieving the highest accuracy (66%) and F1-score (0.587), outperforming state-of-the-art models while maintaining computational efficiency.
- Comprehensive performance evaluations highlighted the impact of input modalities, where edge-based images significantly improved detection accuracy. Models like Swin_T_Edge and Max ViT demonstrated strong generalization across datasets, validating their robustness for real-world applications.
- Statistical analysis using t-tests confirmed the superiority of the Swin_T_Edge model over other models, emphasizing its reliability and potential for practical deployment in wildfire monitoring and disaster management systems.

The following research studies serve as the foundation for this chapter:

❖ Abhishek Verma**,** Virender Ranga, Dinesh Kumar Vishwakarma, "Investigating the Performance vs Computational Complexity Tradeoff in Cross-Domain Fire Risk Detection." in Signal ,Image and Video Processing

This chapter concludes the study on optimizing fire risk detection, summarizing the key contributions and future research avenues.

The next chapter discusses the conclusion, future scope and social impact.

Chapter 8: CONCLUSION, FUTURE SCOPE and SOCIAL IMPACT

This chapter finalizes the research on optimizing cross-domain fire risk detection using advanced vision models. The key contributions of this study are summarized as follows:

- **Enhanced Fire Risk Detection Accuracy**: The proposed Swin_T_Edge model demonstrated superior classification performance with an accuracy of 66% and an F1-score of 0.587, outperforming state-of-the-art models like DenseNet, EfficientNet-B0, and ViT. The model maintained a balance between high accuracy and computational efficiency, highlighting its potential for practical deployment in fire risk monitoring systems.
- **Impact of Input Modalities and Model Generalization**: Through comprehensive performance evaluations, it was observed that edge-based inputs significantly improved model accuracy, particularly for Swin Transformer variants. The models, especially Swin_T_Edge and Max ViT, also demonstrated robust generalization capabilities across diverse datasets, confirming their adaptability in real-world fire detection scenarios.
- **Statistical Validation and Practical Relevance**: The statistical analysis, supported by t-test evaluations, confirmed the superiority of the Swin_T_Edge model over other models with p-values indicating significant performance improvements. This positions the Swin_T_Edge as a reliable solution for wildfire monitoring, disaster management, and environmental conservation.

Future Work

Despite the promising results achieved in this study, several avenues for future research remain open:

- **Integration of Temporal Data**: Incorporating temporal sequences from remote sensing imagery could enhance the dynamic understanding of fire progression, leading to improved prediction accuracy in rapidly changing fire environments.

- **Ensemble Learning and Hybrid Models**: Future work could explore ensemble learning strategies or hybrid architectures that combine the strengths of different vision models to further improve fire risk detection performance and generalization.
- **Utilization of Multispectral and Hyperspectral Data**: Expanding the input modalities to include multispectral or hyperspectral imagery may provide deeper insights into vegetation health and other environmental factors contributing to fire risk, thus enhancing model robustness.
- **Real-Time Deployment and Edge Computing**: Further research should focus on optimizing the models for real-time deployment in resource-constrained environments, utilizing edge computing technologies to enable timely and efficient fire risk monitoring.

Social Impact

The advancements in fire risk detection presented in this study have significant implications for both environmental sustainability and public safety:

- **Wildfire Management and Disaster Response**: The proposed models can be integrated into early warning systems, enabling faster and more accurate identification of high-risk areas, thereby facilitating proactive wildfire management and reducing the devastating impacts on communities and ecosystems.
- **Environmental Conservation and Climate Research**: By improving the accuracy and efficiency of fire risk detection, this research contributes to better management of natural resources and supports climate change mitigation efforts. Accurate fire risk assessments can inform policies aimed at reducing deforestation, protecting biodiversity, and preserving carbon sinks.
- **Public Health and Safety**: The ability to predict and monitor fire risks effectively can help mitigate the health hazards associated with wildfires, such as respiratory issues from smoke inhalation. This research supports the development of tools that can safeguard human lives, property, and infrastructure from wildfire-related disasters.

Overall, this work contributes to the development of intelligent, data-driven fire risk assessment systems, promoting sustainable environmental management and enhancing community resilience to wildfire threats.

References

[1] A. B. Morancho, “A hedonic valuation of urban green areas,” *Landscape and Urban Planning*, vol. 66, no. 1, pp. 35–41, Dec. 2003, doi: 10.1016/S0169-2046(03)00093-8.

[2] “Wildfire in Australia during 2019-2020, Its Impact on Health, Biodiversity and Environment with Some Proposals for Risk Management: A Review.” Accessed: Oct. 28, 2023. [Online]. Available: https://www.scirp.org/journal/paperinformation.aspx?paperid=110099

[3] J. Lelieveld, J. S. Evans, M. Fnais, D. Giannadaki, and A. Pozzer, “The contribution of outdoor air pollution sources to premature mortality on a global scale,” *Nature*, vol. 525, no. 7569, Art. no. 7569, Sep. 2015, doi: 10.1038/nature15371.

[4] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, “Land Use Classification in Remote Sensing Images by Convolutional Neural Networks,” Aug. 01, 2015, *arXiv*: arXiv:1508.00092. Accessed: Oct. 28, 2023. [Online]. Available: http://arxiv.org/abs/1508.00092

[5] F. E. Fassnacht *et al.*, “Review of studies on tree species classification from remotely sensed data,” *Remote Sensing of Environment*, vol. 186, pp. 64–87, Dec. 2016, doi: 10.1016/j.rse.2016.08.013.

[6] S. Liu and Q. Shi, “Local climate zone mapping as remote sensing scene classification using deep learning: A case study of metropolitan China,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 164, pp. 229–242, Jun. 2020, doi: 10.1016/j.isprsjprs.2020.04.008.

[7] R. K. Jaiswal, S. Mukherjee, K. D. Raju, and R. Saxena, “Forest fire risk zone mapping from satellite imagery and GIS,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 4, no. 1, pp. 1–10, Aug. 2002, doi: 10.1016/S0303-2434(02)00006-5.

[8] A. Akay and A. Erdoğan, “GIS-based Multi-criteria Decision Analysis for Forest Fire Risk Mapping,” Oct. 2017.

[9] M. Huesca, J. Litago, A. Palacios-Orueta, F. Montes, A. Sebastián-López, and P. Escribano, “Assessment of forest fire seasonality using MODIS fire potential: A time series approach,” *Agricultural and Forest Meteorology*, vol. 149, no. 11, pp. 1946–1955, Nov. 2009, doi: 10.1016/j.agrformet.2009.06.022.

[10] M. Yadav and L. Das, “Formulation and evaluation of the radius of maximum wind of the tropical cyclones over the North Indian Ocean basin”.

[11] M. Guarnieri and J. R. Balmes, “Outdoor air pollution and asthma,” *The Lancet*, vol. 383, no. 9928, pp. 1581–1592, May 2014, doi: 10.1016/S0140-6736(14)60617-6.

[12] W. Bank and I. for H. M. and Evaluation, “The Cost of Air Pollution: Strengthening the Economic Case for Action,” Sep. 2016, doi: 10.1596/25013.

[13] J. B. van Geffen K. Folkert; Eskes, Henk; Sneep, Maarten; Linden, Mark ter; Zara, Marina; Veefkind, J. Pepijn, “S5P TROPOMI NO2 slant column retrieval : Method, stability, uncertainties and comparisons with OMI,” *Atmospheric Measurement Techniques*, vol. 13, no. 3, pp. 1315–1335, 2020, doi: 10.5194/amt-13-1315-2020.

[14] H. B. Eskes K. F., “Averaging kernels for DOAS total-column satellite retrievals,” *Atmospheric Chemistry and Physics*, vol. 3, no. 5, pp. 1285–1291, 2003, doi: 10.5194/acp-3-1285-2003.

[15] R. V. Martin, “Satellite remote sensing of surface air quality,” *Atmospheric Environment*, vol. 42, no. 34, pp. 7823–7843, 2008, doi: 10.1016/j.atmosenv.2008.07.018.

[16] K. C. Gui Huizheng; Zeng, Zhaoliang; Wang, Yaqiang; Zhai, Shixian; Wang, Zemin; Luo, Ming; Zhang, Lei; Liao, Tingting; Zhao, Hujia; Li, Lei; Zheng, Yu; Zhang, Xiaoye, “Construction of a virtual PM2.5 observation network in China based on high-density surface meteorological observations using the Extreme Gradient Boosting model.,” *Environment international*, vol. 141, no. NA, pp. 105801-NA, 2020, doi: 10.1016/j.envint.2020.105801.

[17] X. Z. Yan Zhou; Luo, Nana; Jiang, Yize; Li, Zhanqing, “New interpretable deep learning model to monitor real-time PM2.5 concentrations from satellite data.,” *Environment international*, vol. 144, no. NA, pp. 106060-NA, 2020, doi: 10.1016/j.envint.2020.106060.

[18] D. N. Thomas and G. Dieckmann, Eds., *Sea ice: an introduction to its physics, chemistry, biology, and geology*. Oxford, UK ; Malden, MA, USA: Blackwell Science, 2003.

[19] P. S. Lee Rick D. ;. McQueen, Jeff, “Air Quality Monitoring and Forecasting,” *Atmosphere*, vol. 9, no. 3, pp. 89-NA, 2018, doi: 10.3390/atmos9030089.

[20] A. Buono, F. Nunziata, and M. Migliaccio, “Analysis of Full and Compact Polarimetric SAR Features Over the Sea Surface,” *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 10, pp. 1527–1531, Oct. 2016, doi: 10.1109/LGRS.2016.2595058.

[21] H. Deng and D. A. Clausi, “Unsupervised segmentation of synthetic aperture Radar sea ice imagery using a novel Markov random field model,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 528–538, Mar. 2005, doi: 10.1109/TGRS.2004.839589.

[22] C. W. Wang Yiyi; Shi, Zhihao; Sun, Jinjin; Gong, Kangjia; Li, Jingyi; Qin, Momei; Wei, Jing; Li, Tiantian; Kan, Haidong; Hu, Jianlin, “Effects of using different exposure data to estimate changes in premature mortality attributable to PM2.5 and O3 in China.,” *Environmental pollution (Barking, Essex : 1987)*, vol. 285, no. NA, pp. 117242–117242, 2021, doi: 10.1016/j.envpol.2021.117242.

[23] C. Varotsos and C. Cartalis, “Re-evaluation of surface ozone over Athens, Greece, for the period 1901–1940,” *Atmospheric Research*, vol. 26, no. 4, pp. 303–310, Aug. 1991, doi: 10.1016/0169-8095(91)90024-Q.

[24] T. S. Li Huanfeng; Yuan, Qiangqiang; Zhang, Xuechen; Zhang, Liangpei, “Estimating ground-level PM2.5 by fusing satellite and station observations: A geo-intelligent deep learning approach,” *Geophysical Research Letters*, vol. 44, no. 23, p. 11,985-11,993, 2017, doi: 10.1002/2017gl075710.

[25] J. Liu, C. Xu, Y. Liu, Z. Zhang, and J. Lu, “Hybrid PM2.5 prediction model based on convolutional neural network and seasonal decomposition method,” *Environmental Science and Pollution Research*, vol. 27, no. 8, pp. 8699–8711, 2020, doi: 10.1007/s11356-019-07594-7.

[26] L. Yao, J. He, J. Sun, F. Gao, and W. Li, “A multi-scale and multi-channel convolutional neural network for PM2.5 prediction,” *Atmospheric Environment*, vol. 244, p. 117962, 2021, doi: 10.1016/j.atmosenv.2020.117962.

[27] B. Z. Zhang Hanwen; Zhao, Gengming; Lian, Jie, “Constructing a PM2.5 concentration prediction model by combining auto-encoder with Bi-LSTM neural networks,” *Environmental Modelling & Software*, vol. 124, no. NA, pp. 104600-NA, 2020, doi: 10.1016/j.envsoft.2019.104600.

[28] Z. Q. Liu Zhulin; Ni, Xiufeng; Dong, Mengting; Ma, Mengying; Xue, Wenbo; Zhang, Qingyu; Wang, Jinnan, "How to apply O3 and PM2.5 collaborative control to practical management in China: A study based on meta-analysis and machine learning.," *The Science of the total environment*, vol. 772, no. NA, pp. 145392–145392, 2021, doi: 10.1016/j.scitotenv.2021.145392.

[29] L. Jian, Y. Zhao, Y.-P. Zhu, M.-B. Zhang, and D. Bertolatti, "An application of ARIMA model to predict submicron particle concentrations from meteorological factors at a busy roadside in Hangzhou, China," *Science of The Total Environment*, vol. 426, pp. 336–345, Jun. 2012, doi: 10.1016/j.scitotenv.2012.03.025.

[30] J. S. Li Xingyang; Sun, Rihui, "A DBN-Based Deep Neural Network Model with Multitask Learning for Online Air Quality Prediction," *Journal of Control Science and Engineering*, vol. 2019, no. NA, pp. 1–9, 2019, doi: 10.1155/2019/5304535.

[31] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[32] S. Agarwal *et al.*, "Air quality forecasting using artificial neural networks with real time dynamic error correction in highly polluted regions," *Science of The Total Environment*, vol. 735, p. 139454, Sep. 2020, doi: 10.1016/j.scitotenv.2020.139454.

[33] L. L. Zhang Dong; Guo, Quansheng, "Deep Learning From Spatio-Temporal Data Using Orthogonal Regularizaion Residual CNN for Air Prediction," *IEEE Access*, vol. 8, no. NA, pp. 66037–66047, 2020, doi: 10.1109/access.2020.2985657.

[34] Z. Shi, Y. Ye, X. Chen, and H. Wang, "Attention-based recurrent neural network for PM2.5 concentration prediction," *Environmental Science and Pollution Research*, vol. 27, no. 22, pp. 27839–27850, 2020, doi: 10.1007/s11356-020-08912-2.

[35] J. B. Wang Lu; Wang, Siqi; Wang, Chen, "Research and application of the hybrid forecasting model based on secondary denoising and multi-objective optimization for air pollution early warning system," *Journal of Cleaner Production*, vol. 234, no. NA, pp. 54–70, 2019, doi: 10.1016/j.jclepro.2019.06.201.

[36] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, Sep. 2021, doi: 10.1016/j.neucom.2021.03.091.

[37] Y. Z. Feng Wenfang; Sun, Dezhi; Zhang, Liqiu, "Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification," *Atmospheric Environment*, vol. 45, no. 11, pp. 1979–1985, 2011, doi: 10.1016/j.atmosenv.2011.01.022.

[38] Y. Y. Wang Qiangqiang; Li, Tongwen; Zhu, Liye; Zhang, Liangpei, "Estimating daily full-coverage near surface O3, CO, and NO2 concentrations at a high spatial resolution over China based on S5P-TROPOMI and GEOS-FP," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 175, no. NA, pp. 311–325, 2021, doi: 10.1016/j.isprsjprs.2021.03.018.

[39] L. Zhu, Y. Hao, Z.-N. Lu, H. Wu, and Q. Ran, "Do economic activities cause air pollution? Evidence from China's major cities," *Sustainable Cities and Society*, vol. 49, p. 101593, Aug. 2019, doi: 10.1016/j.scs.2019.101593.

[40] A. P. Alimissis Kostas; Tzanis, Chris G. ;. Deligiorgi, Despina, "Spatial estimation of urban air pollution with the use of artificial neural network models," *Atmospheric Environment*, vol. 191, no. NA, pp. 205–213, 2018, doi: 10.1016/j.atmosenv.2018.07.058.

[41] C. Gariazzo *et al.*, "A multi-city air pollution population exposure study: Combined use of chemical-transport and random-Forest models with dynamic population data," *Science of The Total Environment*, vol. 724, p. 138102, Jul. 2020, doi: 10.1016/j.scitotenv.2020.138102.

[42] A. Heydari, M. Majidi Nezhad, D. Astiaso Garcia, F. Keynia, and L. De Santoli, "Air pollution forecasting application based on deep learning model and optimization algorithm," *Clean Techn Environ Policy*, vol. 24, no. 2, pp. 607–621, Mar. 2022, doi: 10.1007/s10098-021-02080-5.

[43] Q. Chang, H. Zhang, and Y. Zhao, "Ambient air pollution and daily hospital admissions for respiratory system–related diseases in a heavy polluted city in Northeast China," *Environ Sci Pollut Res*, vol. 27, no. 9, pp. 10055–10064, Mar. 2020, doi: 10.1007/s11356-020-07678-8.

[44] C. Huang, K. Sun, J. Hu, T. Xue, H. Xu, and M. Wang, "Estimating 2013–2019 NO2 exposure with high spatiotemporal resolution in China using an ensemble model," *Environmental Pollution*, vol. 292, p. 118285, Jan. 2022, doi: 10.1016/j.envpol.2021.118285.

[45] J. Douros *et al.*, “Comparing Sentinel-5P TROPOMI NO_2 column observations with the CAMS regional air quality ensemble,” *Geoscientific Model Development*, vol. 16, no. 2, pp. 509–534, Jan. 2023, doi: 10.5194/gmd-16-509-2023.

[46] R. Rakholia, Q. Le, B. Quoc Ho, K. Vu, and R. Simon Carbajo, “Multi-output machine learning model for regional air pollution forecasting in Ho Chi Minh City, Vietnam,” *Environment International*, vol. 173, p. 107848, Mar. 2023, doi: 10.1016/j.envint.2023.107848.

[47] A. Hasnain *et al.*, “Time Series Analysis and Forecasting of Air Pollutants Based on Prophet Forecasting Model in Jiangsu Province, China,” *Frontiers in Environmental Science*, vol. 10, 2022, Accessed: Jan. 03, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fenvs.2022.945628

[48] J. Wei *et al.*, “Ground-Level NO2 Surveillance from Space Across China for High Resolution Using Interpretable Spatiotemporally Weighted Artificial Intelligence,” *Environ. Sci. Technol.*, vol. 56, no. 14, pp. 9988–9998, Jul. 2022, doi: 10.1021/acs.est.2c03834.

[49] E. Marinov, D. Petrova-Antonova, and S. Malinov, “Time Series Forecasting of Air Quality: A Case Study of Sofia City,” *Atmosphere*, vol. 13, no. 5, Art. no. 5, May 2022, doi: 10.3390/atmos13050788.

[50] Y. Guo and Z. Mao, “Long-Term Prediction Model for NOx Emission Based on LSTM–Transformer,” *Electronics*, vol. 12, no. 18, Art. no. 18, Jan. 2023, doi: 10.3390/electronics12183929.

[51] “A dynamic local thresholding technique for sea ice classification | IEEE Conference Publication | IEEE Xplore.” Accessed: Jun. 06, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/322252

[52] T. Zhang *et al.*, “Deep Learning Based Sea Ice Classification with Gaofen-3 Fully Polarimetric SAR Data,” *Remote Sensing*, vol. 13, no. 8, Art. no. 8, Jan. 2021, doi: 10.3390/rs13081452.

[53] Z. Huang, M. Datcu, Z. Pan, and B. Lei, “Deep SAR-Net: Learning objects from signals,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 161, pp. 179–193, Mar. 2020, doi: 10.1016/j.isprsjprs.2020.01.016.

[54] X. Chen, K. A. Scott, M. Jiang, Y. Fang, L. Xu, and D. A. Clausi, “Sea ice classification with dual-polarized SAR imagery: a hierarchical pipeline,” in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, Waikoloa, HI, USA: IEEE, Jan. 2023, pp. 224–232. doi: 10.1109/WACVW58289.2023.00028.

[55] “Remote Sensing | Free Full-Text | Multi-Featured Sea Ice Classification with SAR Image Based on Convolutional Neural Network.” Accessed: Jun. 06, 2024. [Online]. Available: https://www.mdpi.com/2072-4292/15/16/4014#B43-remotesensing-15-04014

[56] Z. Huang, X. Yao, Y. Liu, C. O. Dumitru, M. Datcu, and J. Han, “Physically Explainable CNN for SAR Image Classification,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 190, pp. 25–37, Aug. 2022, doi: 10.1016/j.isprsjprs.2022.05.008.

[57] M. Munsif, H. Afridi, M. Ullah, S. D. Khan, F. Alaya Cheikh, and M. Sajjad, “A Lightweight Convolution Neural Network for Automatic Disasters Recognition,” in *2022 10th European Workshop on Visual Information Processing (EUVIP)*, Sep. 2022, pp. 1–6. doi: 10.1109/EUVIP53989.2022.9922799.

[58] N. Dilshad, S. U. Khan, N. S. Alghamdi, T. Taleb, and J. Song, “Toward Efficient Fire Detection in IoT Environment: A Modified Attention Network and Large-Scale Data Set,” *IEEE Internet of Things Journal*, vol. 11, no. 8, pp. 13467–13481, Apr. 2024, doi: 10.1109/JIOT.2023.3336931.

[59] H. Yar, Z. A. Khan, T. Hussain, and S. W. Baik, “A modified vision transformer architecture with scratch learning capabilities for effective fire detection,” *Expert Systems with Applications*, vol. 252, p. 123935, Oct. 2024, doi: 10.1016/j.eswa.2024.123935.

[60] H. Yar, Z. A. Khan, I. Rida, W. Ullah, M. J. Kim, and S. W. Baik, “An efficient deep learning architecture for effective fire detection in smart surveillance,” *Image and Vision Computing*, vol. 145, p. 104989, May 2024, doi: 10.1016/j.imavis.2024.104989.

[61] A. Khan and M. Sudheer, “Machine learning-based monitoring and modeling for spatio-temporal urban growth of Islamabad,” *The Egyptian Journal of Remote Sensing and Space Science*, vol. 25, no. 2, pp. 541–550, Aug. 2022, doi: 10.1016/j.ejrs.2022.03.012.

[62] Md. N. Mowla, D. Asadi, K. N. Tekeoglu, S. Masum, and K. Rabie, “UAVs-FFDB: A high-resolution dataset for advancing forest fire detection and monitoring using unmanned

aerial vehicles (UAVs)," *Data in Brief*, vol. 55, p. 110706, Aug. 2024, doi: 10.1016/j.dib.2024.110706.

[63] F. Belarbi, A. Hassini, and N. K. Benamara, "A novel approach based on convolutional neural networks ensemble for fire detection," *SIViP*, Aug. 2024, doi: 10.1007/s11760-024-03508-3.

[64] H. Yan, Z. Cui, H. Zhao, J. Zhang, J. Qin, and Q. Guo, "The Multi-Scale Depth-Separable Convolution Network for Fire and Smoke Detection," *Combustion Science and Technology*, vol. 0, no. 0, pp. 1–25, doi: 10.1080/00102202.2024.2372689.

[65] O. H. Kombo, S. Kumaran, E. Ndashimye, and A. Bovim, "An Ensemble Mode Decomposition Combined with SVR-RF Model for Prediction of Groundwater Level: The Case of Eastern Rwandan Aquifers," in *Cybernetics Perspectives in Systems*, R. Silhavy, Ed., in Lecture Notes in Networks and Systems. Cham: Springer International Publishing, 2022, pp. 312–328. doi: 10.1007/978-3-031-09073-8_27.

[66] R. Y. Mintz Brent R. ;. Svrcek, William Y., "Fuzzy logic modeling of surface ozone concentrations," *Computers & Chemical Engineering*, vol. 29, no. 10, pp. 2049–2059, 2005, doi: 10.1016/j.compchemeng.2005.01.008.

[67] P. J. G. L. Nieto F. Sánchez; García-Gonzalo, Esperanza; de Cos Juez, F. J., "PM10 concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM, MLP, VARMA and ARIMA: A case study.," *The Science of the total environment*, vol. 621, no. NA, pp. 753–761, 2017, doi: 10.1016/j.scitotenv.2017.11.291.

[68] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[69] R. Cahuantzi, X. Chen, and S. Güttel, "A comparison of LSTM and GRU networks for learning symbolic sequences," Jan. 04, 2023, *arXiv*: arXiv:2107.02248. doi: 10.48550/arXiv.2107.02248.

[70] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Networks," Apr. 10, 2017, *arXiv*: arXiv:1611.05431. doi: 10.48550/arXiv.1611.05431.

[71] R. Y. Mintz Brent R. ;. Svrcek, William Y., “Fuzzy logic modeling of surface ozone concentrations,” *Computers & Chemical Engineering*, vol. 29, no. 10, pp. 2049–2059, 2005, doi: 10.1016/j.compchemeng.2005.01.008.

[72] E. Žunić, K. Korjenić, K. Hodžić, and D. Đonko, “Application of Facebook’s Prophet Algorithm for Successful Sales Forecasting Based on Real-world Data,” *IJCSIT*, vol. 12, no. 2, pp. 23–36, Apr. 2020, doi: 10.5121/ijcsit.2020.12203.

[73] S. S. Hochreiter Jürgen, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[74] A. Vaswani *et al.*, “Attention is All you Need,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Apr. 30, 2023. [Online]. Available: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[75] I. Tolstikhin *et al.*, “MLP-Mixer: An all-MLP Architecture for Vision,” Jun. 11, 2021, *arXiv*: arXiv:2105.01601. doi: 10.48550/arXiv.2105.01601.

[76] E. Huynh, “Vision Transformers in 2022: An Update on Tiny ImageNet,” May 21, 2022, *arXiv*: arXiv:2205.10660. Accessed: Oct. 28, 2023. [Online]. Available: http://arxiv.org/abs/2205.10660

[77] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” Jan. 28, 2018, *arXiv*: arXiv:1608.06993. doi: 10.48550/arXiv.1608.06993.

[78] M. Ding, B. Xiao, N. Codella, P. Luo, J. Wang, and L. Yuan, “DaViT: Dual Attention Vision Transformers,” in *Computer Vision – ECCV 2022*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds., Cham: Springer Nature Switzerland, 2022, pp. 74–92. doi: 10.1007/978-3-031-20053-3_5.

[79] H. Zhang *et al.*, “ResNeSt: Split-Attention Networks,” presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 2736–2746. Accessed: Dec. 15, 2024. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2022W/ECV/html/Zhang_ResNeSt_Split-Attention_Networks_CVPRW_2022_paper.html

[80] Z. Liu *et al.*, “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” Aug. 17, 2021, *arXiv*: arXiv:2103.14030. doi: 10.48550/arXiv.2103.14030.

[81] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, “Aggregated Residual Transformations for Deep Neural Networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Jul. 2017, pp. 5987–5995. doi: 10.1109/CVPR.2017.634.

[82] Z. Wang, W. Jiang, Y. Zhu, L. Yuan, Y. Song, and W. Liu, “DynaMixer: A Vision MLP Architecture with Dynamic Mixing,” Jun. 18, 2022, *arXiv*: arXiv:2201.12083. doi: 10.48550/arXiv.2201.12083.

[83] S. Shen, S. Seneviratne, X. Wanyan, and M. Kirley, “FireRisk: A Remote Sensing Dataset for Fire Risk Assessment with Benchmarks Using Supervised and Self-supervised Learning,” Sep. 21, 2023, *arXiv*: arXiv:2303.07035. Accessed: Oct. 28, 2023. [Online]. Available: http://arxiv.org/abs/2303.07035

[84] P. Feng and Z. Tang, “A survey of visual neural networks: current trends, challenges and opportunities,” *Multimedia Systems*, vol. 29, no. 2, pp. 693–724, Apr. 2023, doi: 10.1007/s00530-022-01003-8.

[85] J. Wang *et al.*, “Deep High-Resolution Representation Learning for Visual Recognition,” Mar. 13, 2020, *arXiv*: arXiv:1908.07919. doi: 10.48550/arXiv.1908.07919.

[86] “[2204.01697] MaxViT: Multi-Axis Vision Transformer.” Accessed: Nov. 03, 2023. [Online]. Available: https://arxiv.org/abs/2204.01697

[87] V. Goyal, S. Sharma, and B. Garg, “Texture Classification Using ResNet and EfficientNet,” in *Machine Intelligence Techniques for Data Analysis and Signal Processing*, D. S. Sisodia, L. Garg, R. B. Pachori, and M. Tanveer, Eds., Singapore: Springer Nature, 2023, pp. 173–185. doi: 10.1007/978-981-99-0085-5_15.

[88] H. Wu, C. Liang, M. Liu, and Z. Wen, “Optimized HRNet for image semantic segmentation,” *Expert Systems with Applications*, vol. 174, p. 114532, Jul. 2021, doi: 10.1016/j.eswa.2020.114532.

[89] Z. Tu *et al.*, “MaxViT: Multi-Axis Vision Transformer,” Sep. 09, 2022, *arXiv*: arXiv:2204.01697. doi: 10.48550/arXiv.2204.01697.

[90] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” Sep. 11, 2020, *arXiv*: arXiv:1905.11946. doi: 10.48550/arXiv.1905.11946.

[91] A. Howard *et al.*, “Searching for MobileNetV3,” Nov. 20, 2019, *arXiv*: arXiv:1905.02244. doi: 10.48550/arXiv.1905.02244.

[92] S. Nandhini and K. Ashokkumar, “An automatic plant leaf disease identification using DenseNet-121 architecture with a mutation-based henry gas solubility optimization algorithm,” *Neural Comput & Applic*, vol. 34, no. 7, pp. 5513–5534, Apr. 2022, doi: 10.1007/s00521-021-06714-z.

[93] R. A. Perdana, A. M. Arimurthy, and Risnandar, “Remote Sensing Scene Classification using ConvNeXt-Tiny Model with Attention Mechanism and Label Smoothing,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 3, Art. no. 3, Jun. 2024, doi: 10.29207/resti.v8i3.5731.

[94] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” Jun. 03, 2021, *arXiv*: arXiv:2010.11929. doi: 10.48550/arXiv.2010.11929.

[95] Z. Zhang, X. Cui, Q. Zheng, and J. Cao, “Land use classification of remote sensing images based on convolution neural network,” *Arab J Geosci*, vol. 14, no. 4, p. 267, Feb. 2021, doi: 10.1007/s12517-021-06587-5.

[96] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, “Masked Autoencoders Are Scalable Vision Learners,” Dec. 19, 2021, *arXiv*: arXiv:2111.06377. doi: 10.48550/arXiv.2111.06377.

[97] “An Efficient Fire Detection Method Based on Multiscale Feature Extraction, Implicit Deep Supervision and Channel Attention Mechanism.” Accessed: Oct. 03, 2024. [Online]. Available: https://xplorestaging.ieee.org/document/9171455?denied=

[98] H. Yar, T. Hussain, Z. A. Khan, D. Koundal, M. Y. Lee, and S. W. Baik, “[Retracted] Vision Sensor-Based Real-Time Fire Detection in Resource-Constrained IoT Environments,” *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, p. 5195508, 2021, doi: 10.1155/2021/5195508.

PROOF OF PUBLICATIONS

SCIE Journal Paper 1:

- **Abhishek Verma,** Virender Ranga, Dinesh Kumar Vishwakarma, "A novel approach for forecasting PM2.5 pollution in Delhi using CATALYST." Published in Environmental Monitoring and Assessment, Volume 196, Article number 340, (2024), (Pub: Springer).

SPRINGER NATURE Link

Notifications Account

Find a journal Publish with us Track your research Search Cart

Home › Environmental Monitoring and Assessment › Article

A novel approach for forecasting PM2.5 pollution in Delhi using CATALYST

Research | Published: 11 November 2023

Volume 195, article number 1457, (2023) Cite this article

Download PDF Access provided by Delhi Technological University

Environmental Monitoring and Assessment

Aims and scope →

Submit manuscript →

Abhishek Verma, Virender Ranga & Dinesh Kumar Vishwakarma

817 Accesses Explore all metrics →

Use our pre-submission checklist →

Avoid common mistakes on your manuscript.

Abstract

Air pollution is one of the main environmental issues in densely populated urban areas like Delhi. Predictions of the PM2.5 concentration must be accurate for pollution reduction strategies and policy actions to succeed. This research article presents a novel approach for forecasting PM2.5 pollution in Delhi by combining a pre-trained CNN model with a transformer-based model called CATALYST (Convolutional and Transformer model for Air Quality Forecasting). This proposed strategy uses a mixture of the two models. To derive

Sections Figures References

Abstract

Introduction

Literature review

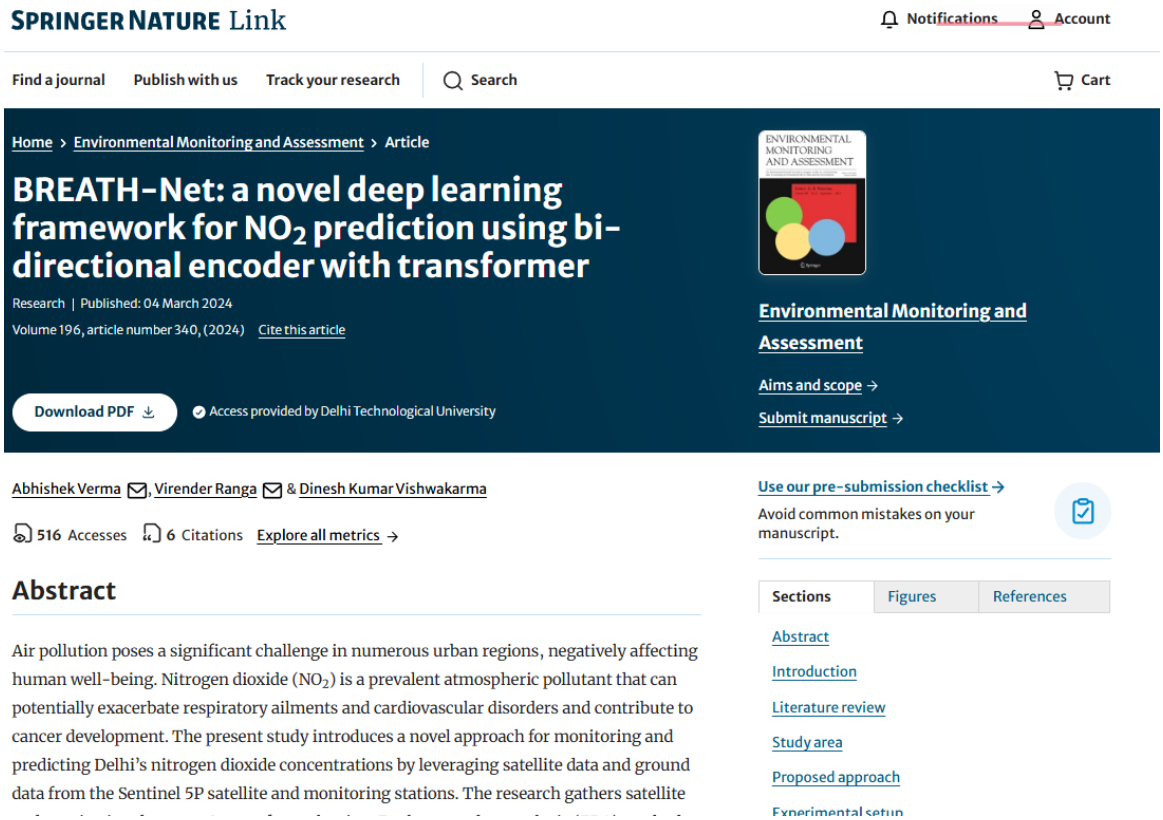
Study area and dataset

Proposed methodology

Results and discussion

SCIE Journal Paper 2:

- **Abhishek Verma,** Virender Ranga, Dinesh Kumar Vishwakarma, "BREATH-Net: a novel deep learning framework for NO_2 prediction using bi-directional encoder with transformer." Published in Environmental Monitoring and Assessment, Volume 196, Article number 340, (2024), (Pub: Springer).

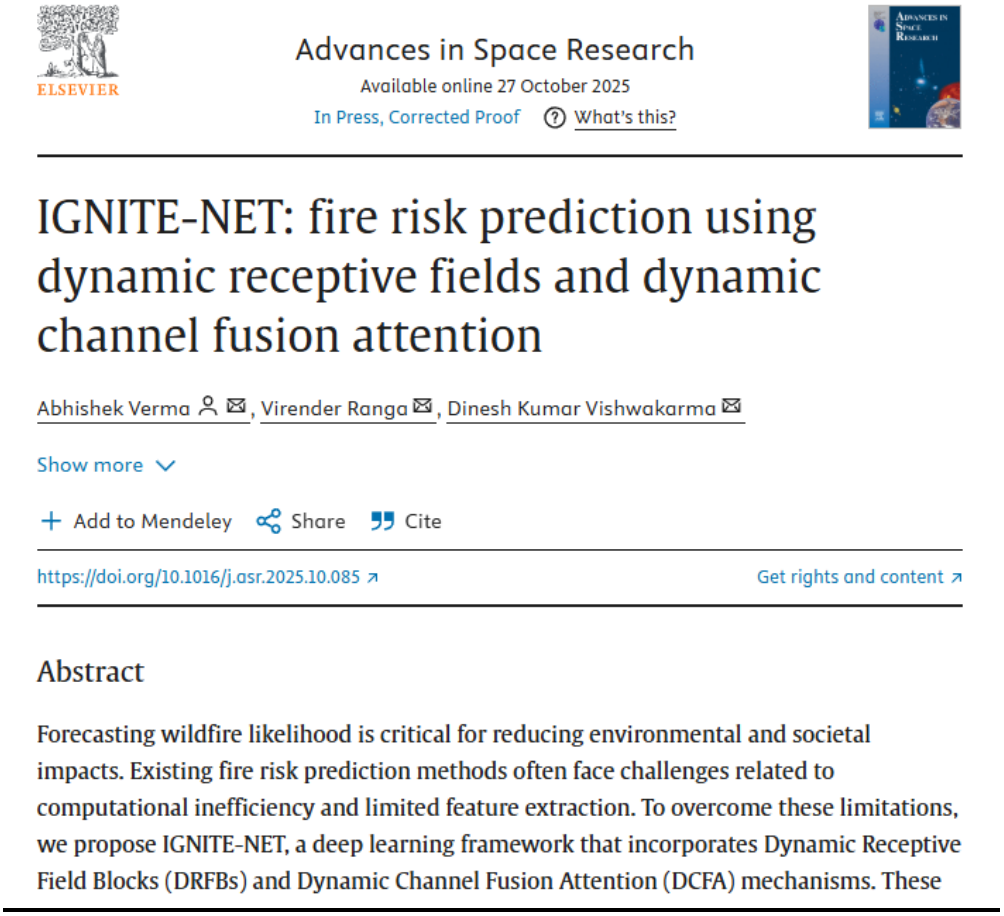


SCIE Journal Paper 3:

- **Abhishek Verma,** Virender Ranga, Dinesh Kumar Vishwakarma, "Investigating the performance vs. computational complexity tradeoff in cross-domain fire risk detection " Published in Signal, Image and Video Processing, Volume 19, page number 713, (2025)(Pub: Springer).

SCIE Journal Paper 4:

- **Abhishek Verma,** Virender Ranga, Dinesh Kumar Vishwakarma, "IGNITE-NET: Fire Risk Prediction using Dynamic Receptive Fields and Dynamic Channel Fusion Attention" Published in Advances in Space Research (2025)(Pub: Elsevier).



ELSEVIER

Advances in Space Research

Available online 27 October 2025

In Press, Corrected Proof What's this?

IGNITE-NET: fire risk prediction using dynamic receptive fields and dynamic channel fusion attention

Abhishek Verma, Virender Ranga, Dinesh Kumar Vishwakarma

Show more

Add to Mendeley Share Cite

https://doi.org/10.1016/j.asr.2025.10.085 Get rights and content

Abstract

Forecasting wildfire likelihood is critical for reducing environmental and societal impacts. Existing fire risk prediction methods often face challenges related to computational inefficiency and limited feature extraction. To overcome these limitations, we propose IGNITE-NET, a deep learning framework that incorporates Dynamic Receptive Field Blocks (DRFBs) and Dynamic Channel Fusion Attention (DCFA) mechanisms. These

Conference Paper 1:

- **Abhishek Verma**, Virender Ranga, Dinesh Kumar Vishwakarma**,** "Forecasting of Satellite Based Carbon-Monoxide Time-Series Data Using a Deep Learning Approach," in 2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT), Institute of Electrical and Electronics Engineers (IEEE), March. 2023, doi: 10.1109/ICITIIT57246.2023.10068609

CERTIFICATE

FOR PAPER PRESENTATION

IEEE

4TH INTERNATIONAL CONFERENCE ON INNOVATIVE TRENDS IN INFORMATION TECHNOLOGY (ICITIIT '23)

IEEE CONFERENCE RECORD NO:57246

This is to certify that ***Prof./Dr./Mr./Ms. Abhishek Verma*** *has presented a paper entitled* ***"Forecasting of Satellite Based Carbon-Monoxide Time-Series Data Using a Deep Learning Approach*** *in the International Conference on Innovative Trends In Information Technology (ICITIIT '23), on 11-12 February 2023, organized by IIIT Kottayam, Kerala, India*

Dr. Kala S
General Co-Chair
ICITIIT '23

Dr. Divya Sindhu Lekha
General Co-Chair
ICITIIT '23

Conference Paper 2:

- **Abhishek Verma**, Virender Ranga, Dinesh Kumar Vishwakarma**,** "Combating Respiratory Health Issues with Intelligent NO_2 Level Prediction from Sentinel 5P Satellite," in 2023 IEEE 20th India Council International Conference (INDICON) at NIT Warangal, Institute of Electrical and Electronics Engineers (IEEE), March. 2023, doi: 10.1109/INDICON59947.2023.10440910

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis: **Design and Development of Ai-Based Framework for Envrionmental and Geospatial Data Analysis**

Total Pages: **139**

Name of the Scholar: **Abhishek Verma**

Supervisors: **Dr. Virender Ranga and Prof. Dinesh Kumar Vishwakarma**

Department: **Information Technology**

This is to report that the above thesis was scanned for similarity detection. The process and outcome are given below:

Software used: Turnitin Similarity Index: 10% Word Count: 33235 Words

Date: 15/02/2025

Candidate's Signature Signature of Supervisors

PLAGIARISM REPORT

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 10 words)

Exclusions

- 5 Excluded Sources

Author Biography

Abhishek Verma is a researcher specializing in Artificial Intelligence and Geospatial Analysis, ear his Ph.D. in Information Technology from Delhi Technological University (DTU), Delhi. The topic of his doctoral dissertation is *"**DESIGN and DEVELOPMENT of AI-BASED FRAMEWORK for ENVIRONMENTAL and GEOSPATIAL DATA ANALYSIS**."* His research focuses on developing AI-driven solutions for environmental monitoring, including wildfire risk detection, air pollution forecasting, and sea ice classification. Abhishek's expertise lies in deep learning, computer vision, and geospatial data analysis, contributing significantly to advancing AI applications in environmental science.