AIR POLLUTION PREDICTION WITH SATELLITE IMAGERY; DEEP LEARNING APPROACH

By

Khumbelo Mulondo (1318840)

Faculty of Science

Research for Submission

Supervisor: Prof Paida Mhangara (Wits University)

A research report submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in partial fulfillment of the requirements for the degree of Master of Science.

September

2024

Declaration

I declare that this research report is my own, unaided work. It is being submitted for the Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

(Signature of candidate)

06th___day of_September___20_24_at_The University of theWitwatersrand

Abstract

This study addresses the urgent global concern of climate change, focusing on air pollution as a significant contributor. Human activities, particularly the burning of fossil fuels for transportation and energy generation, release substantial amounts of harmful air pollutants, including greenhouse gases (GHGs). To effectively mitigate air pollution and its environmental impacts, accurate prediction of its spatio-temporal patterns is crucial.

This study uses the Convolutional Long Short-Term Memory (ConvLSTM) deep learning model to analyze and forecast air pollution using high-resolution, multi-spectral time-series imagery from the Sentinel-2 satellite and meteorological data. By combining the spatial feature extraction capabilities of Convolutional Neural Networks (CNNs) with the sequential data processing strengths of LSTM networks, ConvLSTMs effectively capture both spectral and temporal characteristics inherentin the satellite data. Additionally, temporal meteorological data from air quality monitoring stations is analyzed using LSTM models to enhance the temporal dimension of the analysis. The strategic combination of ConvLSTM for spatial andspectral analysis and LSTM for temporal analysis offers a comprehensive methodology for environmental monitoring and forecasting, with significant implications for urban planning and public health initiatives.

In analyzing nitrogen dioxide spatial distribution, the ConvLSTM model demonstrated consistent improvement in training and validation loss, indicating enhanced generalization capabilities without significant overfitting. However, challenges arise due to the variability in the Structural Similarity Index (SSIM) which was averaging +-0.30 across predicted frames, stemming from the quality variation in the data. The ConvLSTM model performs better with meteorological data, as evidenced by convincing heatmaps and effective avoidance of overfitting where the loss: was 0.1237 and val_loss: of 0.1458 with the 100 epochs. The predicted results are showed in this paper.

Keywords: Greenhouse gases, Air pollution, Climate change, *Machine* Learning, Deep Learning, Convolutional Neural Network, Long Term Short-Term Memory.

Acknowledgements

First and foremost, I express my gratitude to the Almighty, the only wise God and my savior, who has bestowed upon me wisdom, knowledge, and understanding according to the riches of His glory. Without His divine guidance, this endeavor would not have been possible.

I extend my heartfelt appreciation to my loved ones. May this paper stand as a testament to our collective dedication, hard work, and unwavering belief that "all things are possible" with determination and faith.

I express my sincere gratitude to Prof Paida Mhangara, Head of the School of Geography, Archaeology & Environmental Studies, for his invaluable guidance, unwavering support, and constant encouragement throughout this journey. His mentorship has been instrumental in fostering my ability to think independently, and his availability to provide assistance whenever needed has been truly appreciated.

Table of Contents

List of figures

List of Tables

List of equations

List of Abbreviations

CHAPTER 1

Background

Sustainable Development Goal (SDG) number 7 of the United Nations (UN) acknowledges the effect air quality has on the health of human beings. Other SDGs dedicated to cities, energy, and health also have a direct link to pollution of air (Rafaj et al., 2018). The birth of the industrial revolution saw a dramatic increase in atmospheric Green House Gases (GHG'S). The major cause of global warming primarily is the rise in (GHG'S) in the atmosphere, with Carbon dioxide playing a significant role and fundamental aspects that result in the occurrence of intense events of weather and the degradation of natural ecosystems.

Understanding the dynamics of worldwide distribution and the total amount ofatmospheric Carbon dioxide (CO_2) forms the core of scientific forecasts, which in turn helps combat climate change and aiding environmental decision-making. Increasing atmospheric (GHG'S) is the major cause behind the increase in global temperatures since these gases are essential to balance the energy of the Earth. Because of the production of (GHG'S), global warming has become a global issue (Joneset al., 2017). SO_2, CH_4, $PM_{2.5}$, CO_2, O_3, and NO_2 are some of the main (GHG'S) in the atmosphere. As (GHG'S) build up in the atmosphere, a shield builds around the world, trapping heat and preventing it from radiatinginto space (Jones et al., 2017). This shield contributes to increased global warming.

Toxic gaseous pollutants are released into the environment while oil refineries operate. Because of how they operate, oil and gas refineries emit dangerous pollutants like SO_2, Black carbon (BC), particulate matter (PM) and CO. These pollutants contain particles that are baneful effects on the wellbeing of humans, environment, and climate, according to studies by Shikwambana etal. (2020). The existence and advancement of humans have been significantly threatened by carbon emissions, which have also negatively impacted the ecosystem.

The rise of greenhouse gas emissions is the driving factor for more frequent natural disasters such as rising sea levels, droughts, violent storms, and a rocketing increase in insurance claims (Solomon et al., 2007; Griel, 2007). Furthermore, the predicted global temperature for the 21st century is 5–6°C, this is because of industrialization globally which ultimately could result in the decrease of the global domestic product by 5-10%, (Stern, 2007). The average temperature now varies from 0.6 to 0.9 Celsius globally (Shahzad, 2015), up from 0.6 to 0.6C between the 1860s and the 2000s.

From the perspectives of sustainable growth and environmental protection, intensive scientific research has proven to be beneficial to the country. This is because it aids in efficiently reducing greenhouse gas emissions, maintaining economic growth, and minimizing environmental degradation. Predicting carbon emissions can be used to create technical aid for approaches toward sustainable development and initiatives that reduce emissions (Zhao et al., 2022).

According to Kweku (2013), lowering emissions of (GHG'S) will retain an affirmative influence on both the sustainability of human civilization and the assurance of rising labor productivity. The worldwide community is aware that excessive carbon emissions endanger the environment and the sustainable development of society.

According to Andrew (2022), measuring and monitoring Greenhouse gas emissions is a fast- developing field in remote sensing which applies differentkinds of platforms, geographic resolution, techniques, geographic resolutions, and technologies. In the last decade, there is a dramatic advancement in remotesensing technologies which brought the ability to map Greenhouse gas emissions, enabling measurement at different resolutions from meters to kilometers.

The current global climate change accord, which is governed by the United Nations Framework Convention on Climate Change (UNFCCC), is made up of the Paris Accord and the Doha Amended Kyoto Protocol. In addition to being a participant in the IPCC and COP, South Africa is an established member of Sendai Agreement on Disaster RiskReduction. Participation of South Africa in each of these international climate agreements demonstrates its political commitment to lowering greenhouse gas emissions (Shikwambana et al., 2020).

Satellites equipped with sensors regularly monitor Earth, capturing valuable data to track and analyze changes on the planet's surface, including climate variations, land transformations, and alterations in water bodies. TheSentinel-2 program, for example, employs sophisticated sensors to gather high-resolution images that document these changes. The dynamic shifts in climate and human-induced globalconditions pose risks to numerous regions and their populations. By mapping these changes and extracting meaningful insights from the data, it becomes possible to forecast upcoming shifts. Consequently, this predictive capability enables the implementation of measures to mitigate potential threats, safeguarding communities, and ecosystems.

1.2 Problem statement

Numerous methodologies have been developed to analyze sequences of images, with each technique offering different strengths and limitations. Support Vector Machines (SVMs) (Pan et al., 2017) have been applied to identify spatial characteristics, yet they struggle to capture the dynamics over time. The effectiveness of optical flow techniques (Woo et al., 2014) can vary significantly based on chosen parameters, indicating a dependency that can limit their utility. While 2D Convolutional Neural Networks (CNNs) are adept at spatial feature extraction, they fall short in recognizing temporal patterns.

Advances in CNN methodologies (Zhou et al., 2016) have shown promise in extracting both spatial and temporal features, delivering improved outcomes. However, these methods often falter when dealing with long-term dependencies. To address these, sequential models such as RNNs and LSTMs (Pascanu et al., 1997, Graves et al., 2013, Donahue et al., 2015, Cho et al., 2014, Karpathy et al., 2015, Ranzato et al., 2014, Srivastava et al., 2015, Sutskever et al., 2014) were introduced, effectively tackling the challenge of vanishing gradients (Pascanu et al., 2015) and enhancing the model's ability to understand extended sequences. Although sequential models excel in temporal feature extraction, their ability to capture spatial details remains limited. Thus, a comprehensive approach to learning spatio-temporal features (Ranzato et al., 2014, Srivastava et al., 2015) is essential for accurate predictions in scenarios involving high-resolution time-series imagery.

1.3 Study Aim

The work aims to use deep learning and remote sensing methods to track, compare, and forecast the spatiotemporal distribution of greenhousegas emissions in Secunda Mpumalanga.

1.4 Research Objectives

1. To predict the spatial distribution of Nitrogen dioxide (NO_2) using deep learning and satellite imagery.

2. To predict gas emissions in Secunda area with the use ofConvLSTM model and meteorological data.

3. Compare the performance of ConvLSTM model with satellite data vs meteorological data from 2018 to 2022.

1.5 Significance and justification of the study

If we can accurately track the ongoing shifts in climate conditions and map these changes to extract meaningful insights, particularly by using historical time series data (like satellite land maps across various spectral bands) of such climate alterations, then we have the capability to forecast the next occurrences of these changes. This predictive capacity enables us to anticipate and prepare for potential disasters triggered by these swift climate transformations. Essentially, by harnessing past and present climate data, we can create models that predict future environmental scenarios, allowing for preemptive planning and mitigation strategies to combat the adverse effects of rapid climate change.

ConvLSTM leverages its ability to analyze both spatial and temporal aspects of time series imagery, making it a powerful tool for predicting future image frames. This characteristic of ConvLSTM allows it to excel in extracting information from sequences of images over time. Consequently, ConvLSTMs are highly effective models for interpreting time series images and can be employed as predictive models to estimate the next frame in a series.

CHAPTER 2

Literature Review

2.1 Introduction

Human activities such as population expansion, industrial advancement, and actions such as deforestation have increased greenhouse gas (GHG) emissions into the earth's atmosphere. This includes CO2, CH4, and N2O emissions, as defined in National Environmental Management Act (NEMA) White Paper 8. This surge in emissions has created challenges for the climate's capacity to adapt to a persistent increase in the global average temperature, a point elaborated in Section 1 of White Paper 8 by NEMA (2008). To attain goals related to carbon peak or neutrality, it becomes essential to establish a robust monitoring network for tracking GHGs. This entails comprehending our current advancements and identifying necessary measures, as highlighted by Sun et al. (2022).

2.2 Major Sources of air pollution/ gas emissions

As per The Academy of South Africa (2019), the primary factors contributing to air pollution caused by combustion include stationery burning facilities and transportation sources. Various governments assign different levels of importance to these diverse sources. Stationary sources such as mines, businesses, and power plants often have minimal emission regulations.

Due to an unstable power grid, individuals who rely on diesel generators, burn coal, or use other low-quality fuels tend to be the major contributors to pollution. Additionally, biomass fuels are extensively used for household activities in less developed countries, resulting in households significantly impacting air pollution and exposing people to various external factors. Controlled biomass burning sources, which involve the burning of agricultural waste and the clearing of land and forests, have a notable impact on air pollution in developing countries. This practice often leads to uncontrolled biomass burning as homes and other waste contribute to the increase in emissions.

Within both the public and private sectors, vehicles powered by petroleum, such as cars, trucks, and buses, are mobile sources of air pollution. They constitute the primary cause of air pollution in cities. Older vehicles that are inadequately maintained and rely on lower-quality gasoline pose a particularly high risk.

Additionally, the release of air pollutants from ships and airplanes are the main contributors of mobility-related air pollution close to ports and airports.

2.3 Health and Social Impact of gas emissions

GHG emissions are detrimental to human health and welfare. The World Health Organization claims that hot weather makes asthma, respiratory, and cardiovascular conditions worse. Children's respiratory problems increase during heat waves, claim Xu et al. (2012). According to Brockner et al. (2014), approximately 65% of the increased mortality rate associated with air pollution can be attributed to the release of gases from combusting fossil fuel.

Studies suggest that rising emissions from carbon have a major negative impact on the mortality of infants in Africa (Shobande, 2020). Worldwide, air pollution continues to be a severe issue affecting the health of the public. According to the World Health Organization (WHO), around 93% of the world's population is at risk of encountering hazardous levels of air pollution. This issue of air pollution holds significant significance as an environmental challenge, with far-reaching effects on both human well-being and the environment, compounded by the complex challenges of climate change.

2.4 The importance of Monitoring GHG

To tackle climate change effectively, having precise data on greenhouse gas (GHG) emissions is essential. This data should identify the responsible parties, industries, and the quantity of emissions they generate (Boesch et al., 2021). These statistics take critical part in evaluating the effectiveness of various programs, strategies, as well as policies designed to mitigate GHG emissions. By assessing the impact of these initiatives, informed decisions can be taken by policy makers in adjusting the overall efficiency to improve their overall efficacy in reducing GHG emissions. Furthermore, it makes it possible for GHG trading systems to function, as these plans would not be permitted without trustworthy trading units. Based on how credible they appear to buyers and sellers, these trading strategies are trusted (Boesch et al., 2021).

Accurate GHG statistics are also needed to investigate how GHGs, and global warming are related (Wunch et al., 2010, 2011). By precisely measuring the diurnal, monthly, seasonal, and inter-annual oscillations of these gases, we may deduce the sources and sinks of significant GHGs. We can understand how the atmosphere and climate interact with anthropogenic and naturally occurring GHG emissions (Wunch et al., 2011). In addition, we can pinpoint the physical and chemical mechanisms behind these variations and project future patterns. Chemical transport models (CTMs) and methods for compiling emission inventories, as well as advancing research into the carbon cycle, would all be made possible by a reliable GHG monitoring system (MacFaul, 2007, Z. D. Yang et al.,2020)

2.5 The use of machine learning for monitoring and predictive analysis of GHG's

Problems with air quality forecasting have frequently been addressed using traditional statistical methods. The idea of utilizing past data to promote learning is prominently utilized in these techniques. ARMA (Bartholomew DJ, 1971) and ARIMA (Kumar et al., 2010) are two important statistical techniques that have been used to forecast air quality. However, due to the extended training time, these approaches are unable to keep up with the real demand as data complexity and volume rise (Bekkar and others, 2021).

Machine learning-based prediction techniques are becoming more common as big data and artificial intelligence develop. Given that these models do not have to understand the chemical or physical attributes of air pollutants (Bekkar et al., 2021). Additionally, because air pollution data are typically nonlinear, the linear assumption restricts these models' ability to predict consequences.

Recurrent neural networks (RNNs), which are known for rapidly collecting vast associations in sequential data, have emerged as the preferred method for handling temporal dependencies. The rise of big data and improvements in artificial intelligence have increased the prominence of machine learning- based prediction approaches. The research has primarily focused on a specific set of machine learning techniques, that consist of logistic regression, multilinear regression (MLR), support vector regression (SVR), support vector machines (SVMs), decision trees (as discussed by Wei et al. in 2019), hidden Markov models (as presented by Chen et al. in 2020), and backpropagation neural networks (as highlighted by Chan). These models have gotten a lot of attention because of their ability to provide findings even without a thorough grasp of the underlying physical or chemical properties of air contaminants.

Basic machine learning (ML) methods are good at finding simple patterns where things do not change in straightforward ways, but they might miss more complex patterns or trends over a long time. Recurrent Neural Networks (RNNs) are a type of advanced tool that's really good at understanding data that changes over time, capturing longer and more complex patterns.

When studying how air pollution levels relate to the weather, researchers often use several different tools. Multiple Linear Regression (MLR) is like trying to draw a straight line through points on a graph to see the general trend. Random Forest (RF), mentioned by Yu and colleagues in 2016, is like using a lot of different lines from different angles to make predictions—it is good for when the data is mixed and messy. Support Vector Regression (SVR), talked about by Lin and others in 2011, is useful when the relationship between data points is complicated and not straight.

Artificial Neural Networks (ANN), as Wang and his team described in 2015, are like building a brain model to understand very complex patterns that the other methods might not catch. Each method has its strengths. MLR is simple; RF is great for dealing with different types of data and finding out what is important; SVR works well when things are too complex for simple lines; and ANNs are powerful for figuring out complicated relationships, making them all useful for studying how air pollution and the weather are connected. These methods can capture non-linear features from raw data to a certain extent, but complex spatio-temporal correlations in the historical data cannot be fully extracted (Yan et al., 2021; Zhang et al., 2020a).

2.6 Advancement of deep learning

Atmospheric pollution needs detailed quantitative ground measurements (point-based) but at the same time a robust spatial and temporal modelling as well in order to detect the most affected areas in time (day, month, season) and to help decision makers to search for a sustainable solution for environmental quality and population health status improvement (Xing j et al., 2015 & Janssens-Maenhout et al., 2015). Following this direction, there is rich international literature focusing on developing digital models of NO2 and for other aerosol-related pollution such as CO or PM2.5, by integrating geospatial data interpolation techniques starting from the existing ground station measurements (Robinson D et al., 2015 & Vardoulakis s et al., 2011), combined with pollutants' dispersion modelling (Beckerman et al., 2013 & Beelen et al., 2013), land use derived geographical regression for cities and regions , and field.

With the rapid growth of artificial intelligence and big data techniques, deep learning (DL) technologies have been extremely used in big data analysis to solve various classification and regression problems, including: computer vision (Bengio et al., 2009), image classification (Chan et al., 2015), speech recognition (Mohamed et al., 2011), policy analysis (Strubell et al., 2020), time series prediction (Zhang et al., 2015) and natural language processing (Collobert and Weston, 2008), especially in the field of air pollution prediction (Zhao et al., 2020; Krishna Rani Samal et al., 2021a; Zhang et al., 2022a). They can not only predict specific values of pollutant concentrations, but also classify pollutant levels as well as achieve long-term prediction tasks for multiple days in the future (Krishna Rani Samal et al., 2021b; Macatangay and Hernandez, 2020). In addition to this, some researchers have shown that deep learning technologies determine the underlying reasons of meteorological conditions, seasonal variations, regional geographical relationships, chemical mechanisms, etc. that can change air pollutant concentrations (Zhang et al., 2021a; Kurt and Oktay, 2010; Bai et al., 2019a; Abdul-Wahab, 2001).

The advantages of deep learning technology in dealing with non-linear spatio-temporal correlations have been gradually discovered, which has dramatically improved the accuracy of pollutants concentration prediction (Qi et al., 2019; Zhou et al., 2022). A typical deep learning method uses a hierarchical model to map the input to the final output (Goodfellow et al., 2016; Reichstein et al., 2019; LeCun et al., 2015; Bartlett et al., 2021). Its modeling idea is to construct certain blocks or layers together to form an end-to-end deep network structure.

Deep learning models are quite suitable for processing historical air pollution data with spatio-temporal correlations. For instance, Convolutional Neural Networks (CNNs) (Rijal et al., 2018; Zhang et al., 2016b; Bo et al., 2018; Chakma et al., 2017) are used to extract spatial correlation in regional multi-site data; Graph Convolutional Neural Networks (GCNs) (Qi et al., 2019; Bruna et al.,

2013) apply convolution operations to graph structure data, thus they can well represent the network of air quality monitoring sites in non-Euclidean space. Recurrent Neural Networks (RNNs), like the Original Recurrent Neural Network (Original RNN) (Elman, 1991; Athira et al., 2018; Singh et al., 2012; Loy-Benitez et al., 2019) and its variants, Gated Recurrent Units (GRUs) (Loy-Benitez et al., 2019; Cho et al., 2014; Huang et al., 2021b; Xu et al., 2021), Long Short-Term Memory (LSTM) (Loy-Benitez et al., 2019; Hochreiter and Schmidhuber, 1997; Lu et al., 2021a; Ulpiani et al., 2022; Zhao et al., 2019; Navares and Aznarte, 2020; Wu et al., 2022), Read-first LSTM (RLSTM) (Zhang et al., 2021a), Long Short-Term Memory Neural Network Extended (LSTME) (Li et al., 2017), can be used to extract temporal correlation in raw data. Convolutional Neural Network- Long Short-Term Memory (CNN-LSTM) based on CNN and LSTM (Pak et al., 2018, 2020; Zhu et al., 2021; Yan et al., 2021; Mengara et al., 2020; Soh et al., 2018; Wen et al., 2019) can be used to extract spatio-temporal correlations from historical data.

Recurrent Neural Networks (RNNs) are designed to understand patterns over time, making them great for dealing with data that changes or follows a sequence. They have become a go-to choose for analysing anything that evolves over time because of their ability to remember previous information and use it to make predictions about what comes next.

However, RNNs can struggle with making predictions far into the future due to problems called gradient vanishing and explosion, which basically means they start forgetting the earlier data or overemphasizing it.

To solve these issues, researchers have developed special versions of RNNs, like LSTM networks and (GRUs). These models, as mentioned by Khatibi and colleagues in 2021, are better at remembering and using information from both recent and distant past, making them more reliable for predicting what happens next over longer periods. They do this by using special gates that regulate the flow of information, keeping what is useful and discarding what is not, which helps avoid the forgetting problem or the overemphasis issue seen in standard RNNs.

In recent research, deep learning techniques have made significant strides in areas like image classification and feature extraction, which are crucial in pattern recognition and computer vision fields. These advancements include developments in object detection and object tracking. The Support Vector Machine (SVM) method, for instance, has been effectively used to pull out spatial features that are vital for classifying images. Meanwhile, the Stacked Autoencoder (SAE) approach has gained attention for its ability to capture both spatial and spectral features, enriching the quality of image classification.

2.7 The application of deep learning and Machine leaning in spatial and temporal data features

Convolutional Neural Networks (CNNs) stand out for their ability to meticulously extract image features, playing a pivotal role in both identifying and classifying objects within images. Over time, various iterations of CNNs have been introduced to refine the process of image segmentation and classification. Despite these advancements, a common limitation among many CNN models is their struggle to effectively capture temporal features - the changes that occur with time - which are as critical as spatial features for comprehensively understanding and analyzing images, especially in dynamic environments where objects or scenes evolve over time.

To effectively capture temporal features, researchers have utilized various sequence models such as RNN, LSTM, and GRU (Junyoung et al., 2014). In (Sutskever et al., 2014), a sequence-to-sequence LSTM encoder-decoder model is employed, training two temporally concatenated LSTMs—one for processing the input sequence and another for generating the output sequence. Additionally, an RNN- based model (Cho et al., 2014) is developed to predict the next video frame by quantizing image patches for interpolating intermediate frames, albeit limited to predicting only one frame ahead. Expanding on this work, (Srivastava et al., 2015) utilizes an LSTM encoder-decoder predictor model to forecast a sequence of images. While this model can reconstruct the input sequence and predict future sequences, it falls short in capturing spatial correlation.

Historically, various models have been suggested for extracting features from sequences of images. The Support Vector Machine (SVM) (Pan et al., 2017) has been utilized for its ability to extract spatial features, though it falls short in capturing temporal dynamics. Optical flow methods (Woo et al., 2014) have also been employed, but their effectiveness heavily depends on the choice of parameters.

Similarly, 2D Convolutional Neural Networks (CNNs) have been adept at spatial feature extraction but lack the capability to understand temporal changes. Several CNN techniques (Zhou et al., 2016) have been developed to address spatio-temporal features, achieving satisfactory results. However, these models often struggle with long-range dependencies, making them less suitable for analyzing data over extended periods.

To overcome these limitations, sequential models such as RNN and LSTM (Pascanu et al., 1997, Graves et al., 2013, Donahue et al., 2015, Cho et al., 2014, Karpathy et al., 2015, Ranzato et al., 2014, Srivastava et al., 2015, Sutskever et al., 2014) have been introduced, specifically targeting the vanishing gradient problem (Pascanu et al., 1997) and thereby effectively managing long-range dependencies (Hochreiter et al., 1997). These sequential models excel at extracting temporal features but traditionally lack the ability to capture spatial details. Therefore, the integration of spatial and temporal learning is essential for accurate predictions, especially in applications involving time-series analysis of high-resolution images. This holistic approach to learning both spatio-temporal features is crucial for enhancing model performance in complex predictive tasks.

Pak et al (2018), used a spatio-temporal model (CNN-LSTM) to predict the average PM2.5 concentration of Beijing in the next day. They introduced the mutual information (MI) estimator to analyze the temporal and spatial correlations and constructed the spatio-temporal feature vector to reflect the non-linear spatio-temporal correlations. Then, they deployed a CNN to extract important spatial correlation and LSTM to process the temporal correlation. The outputs of the two modules were regard as spatio-temporal correlations for final prediction (Pak et al., 2020). Another similar hybrid deep learning model is for the non-Euclidean structure data. Qi et al. developed a model of Graph Convolutional Neural Network-Long Short-Term Memory (GCN-LSTM) to predict air pollutant concentration and obtained better results, compared to CNN-LSTM model (Qi et al., 2019).

In recent years, advancements in machine learning, particularly in deep learning, have made it possible to address the challenge of simultaneously capturing spatial and temporal features in image sequences. By developing a well-structured end-to- end system (see figure 1) and leveraging sufficient data, this issue can be effectively resolved by integrating convolutional operations with LSTM (Shi et al., 2015, Mendel et al., 2016, Chong et al., 2017). The use of ConvLSTM (Shi et al., 2015) represents a significant step forward, enabling the extraction of both spatial and temporal features from images to predict subsequent frames accurately. This paper explores one variant ConvLSTM model with two different data types. This model is trained end-to-end to predict the outcome images.

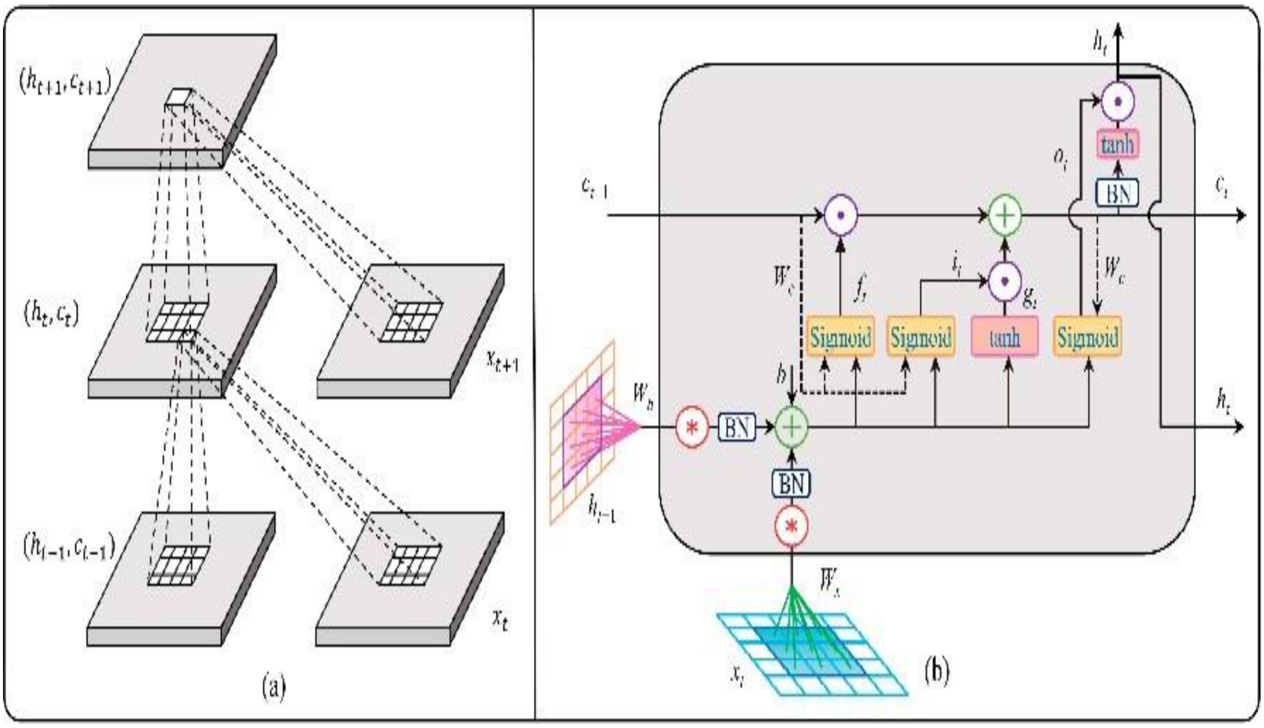

Figure 1: ConvLSTM operational layout

Chapter 3

Methodology

3.1 Study area of Mpumalanga.

Sasol owns the Secunda CTL synthetic fuel facility in Secunda, Mpumalanga, South Africa. The method the business employs to create synthetic crude oil that resembles petroleum is coal liquefaction. The Fischer-Tropsch process regulates the dynamics of this process. It is both the largest coal liquefaction operation and the single largest producer of (GHG'S) in the cordiality ([Secunda CTL - Wikipedia]).

The Sasol II unit, which began operations in the early 1980s, and the Sasol III unit, which started operations in 1984, were the two production units in the second CTL. The factory can produce roughly 160,000 barrels (25,000 m3/d) per day. The Sasol III Steam Plant, which features a 301 m (988 ft) tall chimney constructed by Concor, is one of the highest structures in Africa. The chimney is made up of four 300 m (980 ft) reinforced concrete flues, a windshield that is 292 m (958 ft) high, and a temporary roof that is 1 m (3.3 ft) high atop the fourth flue.

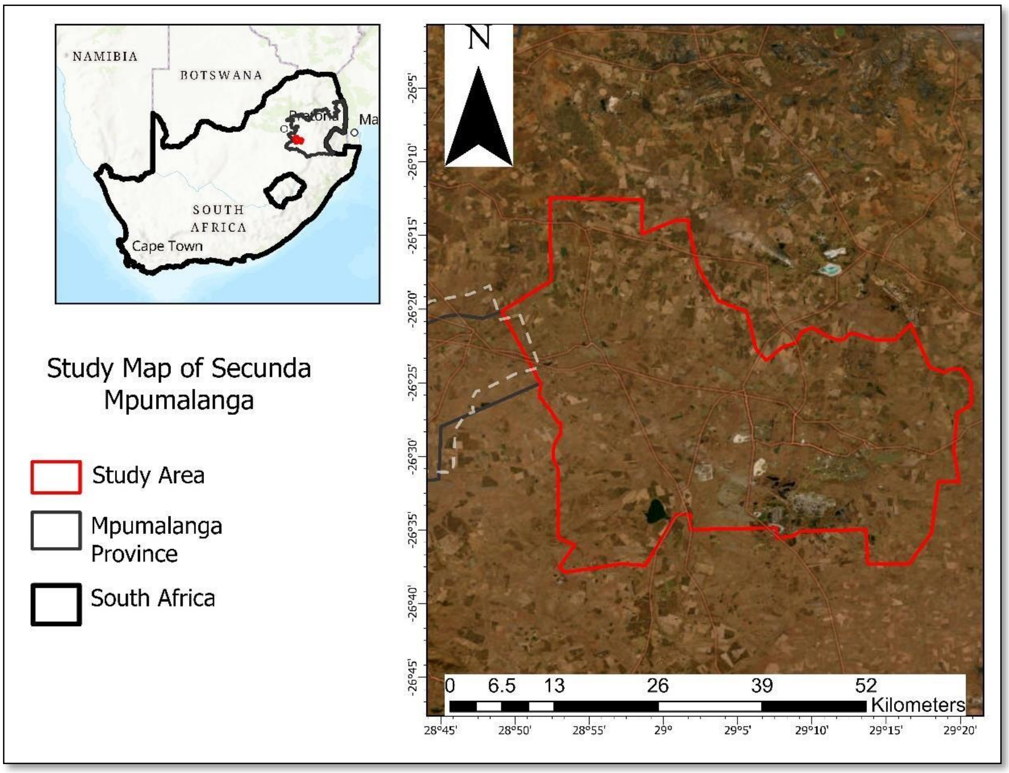

Figure 2:Study Map for South Africa's Sasol secunda plant as displayed by ArcGIS Pro

3.2 Data Collection

3.2.1 Satellite Imagery

The data originates from the Sentinel-2 satellite, part of the Copernicus program by the European Union. Launched on June 23, 2015, Sentinel-2 orbits the Earth at an altitude of 786 kilometers and completes approximately 14.3 revolutions daily. It is equipped with the Multispectral Instrument (MSI), which captures data across 13 spectral bands. These bands encompass a range from Visible and Near-Infrared (VNIR) to Shortwave Infrared (SWIR) wavelengths, covering a swath width of 290 kilometers. This broad spectral coverage and high revisit rate make Sentinel- 2 an invaluable asset for Earth observation, particularly in applications such as monitoring vegetation, land use and land cover changes, and environmental management.

The Multispectral Instrument (MSI) on the Sentinel-2 satellite measures the radiance reflected from Earth's surface and atmosphere across 13 spectral bands with varying spatial resolutions tailored to different observational needs: 4 bands at a 10-meter resolution are designed for high-detail observation. (See figure 3 as input image). These include:

- Blue (490 nm)
- Green (560 nm)
- Red (665 nm)
- Near-Infrared (NIR) (842 nm)

These bands are essential for capturing detailed imagery that can be used for monitoring vegetation health, urban planning, and coastal waters. 6 bands at a 20-meter resolution provide a balance between detail and coverage, including: Four narrow bands optimized for vegetation characterization (705 nm, 740 nm, 783 nm, 865 nm) Two larger Shortwave Infrared (SWIR) bands (1,610 nm and 2,190 nm) that are crucial for detecting snow, ice, clouds, or assessing vegetation moisture stress. The 20-meter bands offer detailed insights into vegetation health and stress levels, as well as the presence of snow, ice, and clouds, supporting agricultural monitoring and environmental studies. 3 bands at a 60-meter resolution are used primarily for atmospheric corrections and cloud screening, including:

- Aerosols (443 nm)
- Water vapor (945 nm)

- Cirrus clouds (1375 nm)

These bands are vital for ensuring the accuracy of observations by correcting for atmospheric interference, thereby improving the quality of data for analysis. This diversified spectral and spatial resolution setup allows Sentinel-2 to offer comprehensive monitoring capabilities across various applications, from precise agricultural mapping to effective management of natural resources and environmental monitoring.

The limitations of Sentinel-2 NO2 retrievals must be recognized. Study by Daniel et al (2023), found that the retrievals are most successful over bright, uniform surfaces. Over these surfaces, a source rate detection limit of about 500 kg NOx h−1 can be inferred from the consistent detection of emissions from Riyadh power plant 9, for which our lowest source rate estimate was 370 kg h−1. Identifying a good reference scene to remove albedo-related artifacts is more challenging for complex surfaces.

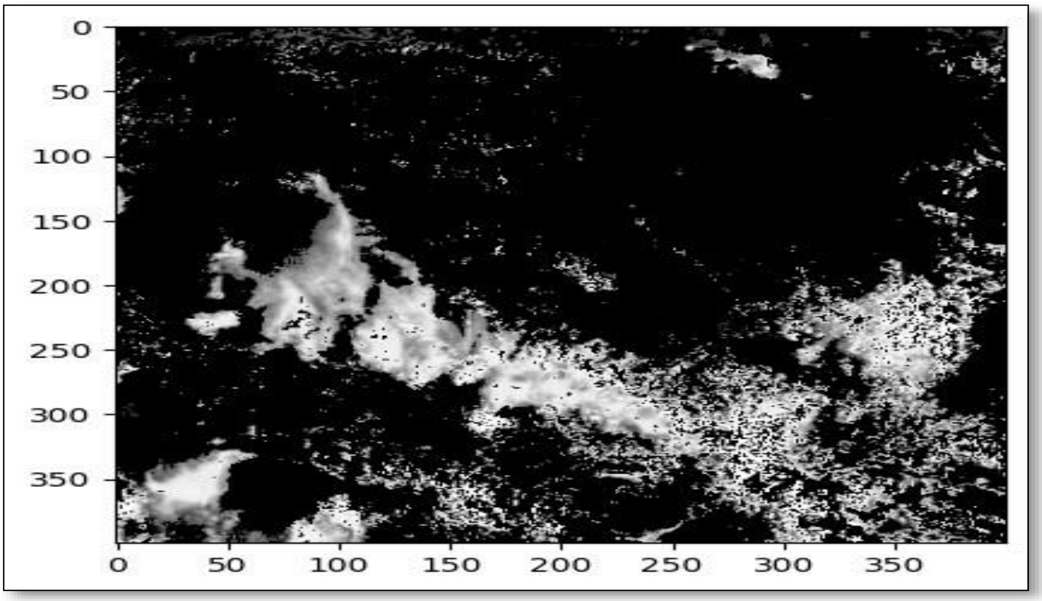

Figure 3: Satellite Imagery with filtered clouds and nitrogen dioxide

3.2.2 Air Quality data for ConvLSTM

The meteorological data utilized in this study was obtained from the South African Air Quality Information System (SAAQIS), which features a network of air monitoring stations located throughout Secunda, Mpumalanga for a period of 5 years (2018-2023). Notably, some of these monitoring stations are owned and operated by Sasol, a major industrial entity. These stations have been strategically positioned around the refinery station to collect data that reflects the impact of industrial activities on air quality. This setup enables a comprehensive analysis of air quality in the vicinity, offering insights into the environmental influences of refinery operations and other industrial activities on the local atmosphere.

A total of five (5) datapoints referring to air quality stations that coverers a radius of 45km were utilized. Each station has entries of more than 30 thousand data inputs. For missing data points due to station's malfunction, we use total average due to its complexity to implement and it helps maintain overall distribution and it has less impact on the correlation between values.

The air monitoring stations are;

- Secunda
- Club_NAQI
- Bosjesspruit_All
- eMbalenhle_North
- Embalenhle

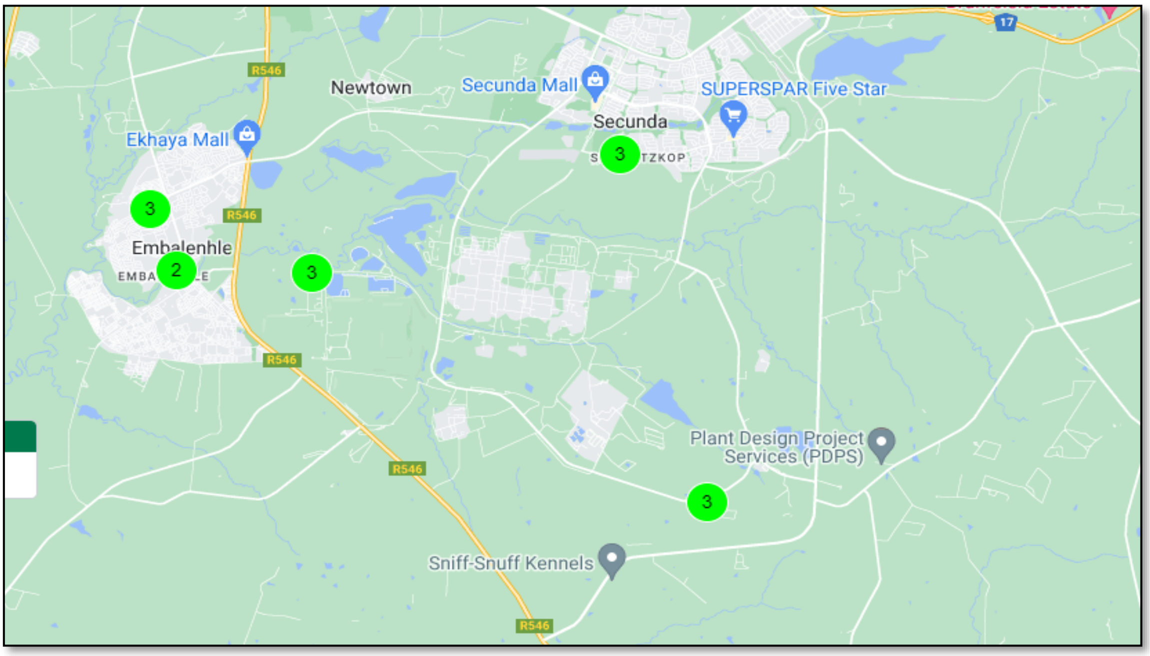

Figure 4: Air Quality Stations from SAWIS

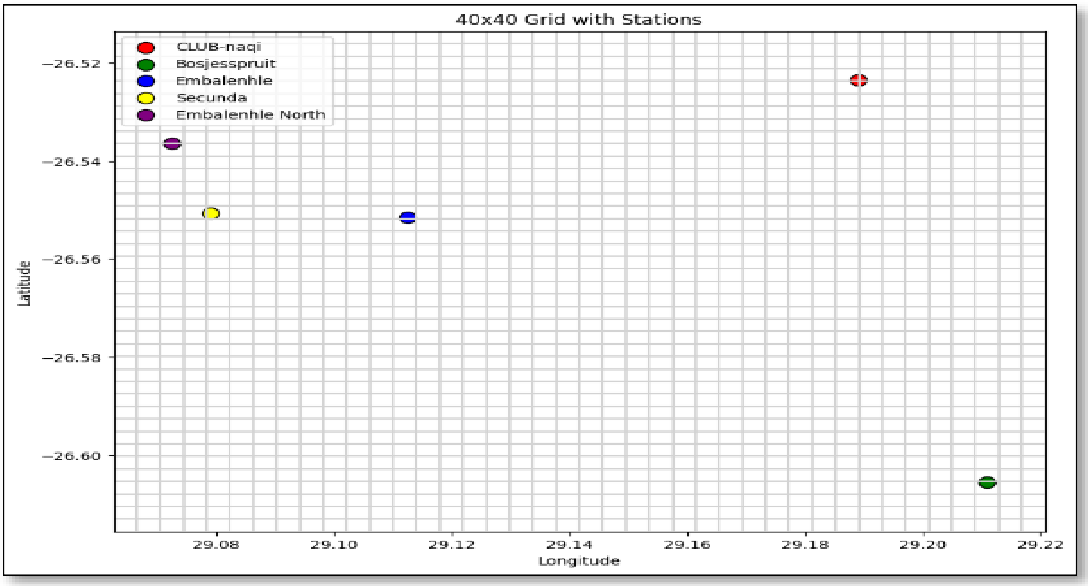

Figure 5: Grid locations of weather stations

Air quality data, typically collected by fixed-ground monitoring stations and various other monitoring systems (Hochreiter et al., 1997, Junyoung et al., 2014), are recorded at set intervals. This method, while reliable, does not effectively capture the dynamic environmental changes occurring in adjacent areas over time (Junyoung et al., 2014). This limitation is particularly pronounced when compared to tasks such as clustering (Zhu et al., 2016), where the data's spatial aspects may be more static. The challenge of making long-term predictions about air quality is significantly compounded by the data's complex non-linear characteristics across both spatial and temporal dimensions. Air pollution levels can change dramatically over time, often fluctuating wildly due to unpredictable events, such as accidents.

Understanding and predicting these dynamic, complex relationships require a novel approach to spatio-temporal data mining. Developing a model that can accurately forecast air quality while accounting for these rapid and significant changes poses a considerable challenge. The key lies in crafting a method that not only encompasses the intricate non-linear interactions within the data but also adapts to the abrupt changes that can occur in the environment, thereby providing accurate, long-term air quality predictions.

3.3 Modeling

3.3.1 Convolution for extraction of spatial features

Convolution basically is a mathematical operation used in Convolutional Neural Networks (CNNs) to perform an affine transformation on input data, such as images, to extract essential features and map these features onto output volumes. In the context of CNNs, the convolution process involves sliding a filter or kernel over the input image and performing element-wise multiplication followed by a sum. This technique is crucial for capturing the vital characteristics of the input images as the data progresses through the network's deeper layers.

The primary purpose of convolution in CNNs is to encode the spatial features of an image. It does this by preserving the relationship between pixels through affine transformations, enabling the network to learn the spatial hierarchies of features. From edges and textures in the initial layers to more complex patterns in the deeper layers, convolution allows CNNs to effectively understand and interpret the spatial structure of images, making it a fundamental operation for tasks involving image recognition, classification, and processing.

The convolution operation in Convolutional Neural Networks (CNNs) typically encompasses both a convolution layer and a pooling layer, with both operations involving affine transformations to process the input data effectively.

Convolution Layer: This layer uses a filter or kernel to perform element-wise dot products that are between the filter and segments of the input image. As the filter slides over the image, it computes the dot product at each position, producing a feature map. This process extracts important features such as edges, shapes, or textures from the image, resulting in another image volume, often referred to as a feature map. The convolution layer is crucial for identifying the various spatial features present in the input image, making it essential for the network's ability to learn complex patterns as data moves through successive layers.

Pooling Layer: Following the convolution layer, the pooling layer further processes the feature map to reduce its spatial dimensions through down sampling operations like max pooling or average pooling. This step reduces the computational complexity for subsequent layers, helps in achieving translation invariance, and prevents overfitting by abstracting higher-level features from the pooled feature maps.

Both the convolution and pooling layers work in tandem within a CNN to efficiently extract and condense the spatial features of an image, preparing the network for tasks such as classification or detection by progressively reducing the spatial size of the representation to focus on the most salient features. Suppose we have image of size (m × n × nc) where m is height, n is width and nc depth in image and a filter f of size (f x f x nc) then output produce by this convolution operation is (m − f + 1 × n − f + 1 × nc).

m: The height of the input image.
n: The width of the input image.
f: The size (height and width) of the filter or kernel used in the convolutional layer.
nc: The number of filters (or channels) used in the convolutional layer.

Equation 1: Convolution layer

Pooling is a technique commonly used in convolutional neural networks (CNNs) to decrease the dimensionality of feature maps generated by convolutional layers. This reduction in size helps to accelerate computation while maintaining important spatial features. Additionally, pooling aids in enhancing feature robustness by focusing on the most significant information within each region of the feature map. By selecting the most relevant features and discarding redundant or less informative ones, pooling contributes to the efficiency and effectiveness of the overall network architecture.

$$\left(\left\lfloor\frac{n+2p-f}{s}\right\rfloor+1\right)\times\left\lfloor\left(\left\lfloor\frac{n+2p-f}{s}\right\rfloor+1\right)\times n_c\right\rfloor .$$

n: The size (height or width) of the input image (assuming a square image for simplicity).
p: The padding applied to the input image.
f: The size (height and width) of the filter or kernel used in the convolutional layer.
s: The stride, which is the number of pixels by which the filter is moved across the input image.
nc: The number of filters (or channels) used in the convolutional layer.

$\lfloor \cdot \rfloor$: The floor function, which rounds down to the nearest integer.

Transposed convolution, often referred to as up-convolution, operates inversely to the standard convolution process found in CNNs. While typical convolutional operations reduce the spatial dimensions of the input data through their filtering process, transposed convolution aims to do the reverse. It takes what could be considered the output of a standard convolution and transforms it back to the size of the convolution's original input. This process is crucial in tasks like image generation or segmentation, where it is necessary to up sample or increase the resolution of a feature map to produce an output image with dimensions similar to the original input. By maintaining a connectivity pattern that matches the corresponding convolution, transposed convolution effectively acts as a 'convolutional decoder,' reconstructing or up sampling data (Chhatra Pratap Bharti., 2022).

The limitation with CNNs is their focus on spatial feature extraction, leaving the temporal dynamics within data unaddressed. Temporal features, which describe how data changes over time, are essential for understanding and predicting sequences, whether in video frames, time-series data, or any other domain where the sequential order and timing of events matter. To capture these temporal dependencies, integrating a sequential model, such as an RNN, LSTM, or Gated Recurrent Unit (GRU), is necessary. These models are designed to process data sequentially, allowing them to learn and remember patterns over time, making them ideal complements to CNNs for tasks that require understanding both spatial and temporal aspects of data (Chhatra Pratap Bharti., 2022).

(LSTM - breakdown)

LSTM networks are a specialized variant of RNNs designed to address and overcome the vanishing gradient problem commonly encountered in standard RNNs. Central to the LSTM architecture is a unique component known as the memory cell (c_t), which serves as the core of the LSTM's ability to preserve information over extended periods (Chhatra Pratap Bharti., 2022).

This memory cell contains several gates — specifically, the input gate, forget gate, and output gate — that regulate the flow of information. These gates decide which data is important to keep, which to discard, and how to use this data to generate the network's output at each step:

Input Gate (it) : When the input gate is activated, it signifies the addition of new information to the memory cell. This gate filters incoming data, deciding what new information is relevant to be stored in the cell state.

Forget Gate (ft) : Activation of the forget gate indicates that some information from the previous cell state ($ct-1$) is deemed irrelevant and can be removed or "forgotten." This mechanism allows the LSTM to discard data that is no longer necessary for future computations, preventing the network from becoming cluttered with outdated information.

Output Gate (ot) : When the output gate is activated, it determines which part of the current cell state (ct) will be passed on to the final state (ht). This gate filters the information from the cell state, deciding what will be used to generate the output of the LSTM at the current timestep (Chhatra Pratap Bharti., 2022).

By selectively filtering information in this manner, LSTMs effectively trap and preserve gradients, thereby mitigating the vanishing gradient problem. This capability allows LSTMs to learn and maintain long-range dependencies within the data, making them particularly adept at tasks involving sequential data where past information is crucial for understanding future outcomes (Chhatra Pratap Bharti., 2022).

According to Chhatra Bharti(2022), the memory cell in a (LSTM) network plays a crucial role in managing information flow across the sequence of data. It operates through the coordinated activity of three distinct gates: the input gate, the forget gate, and the output gate. Each gate has a specific function, influenced by the information from the previous cell state (1ct−1), and collectively, they determine how information is updated and transferred through the network.. The key equations of LSTM are shown below, where '∘' denotes the Hadamard product:

Equation 3:LSTM Model

$$ft = \sigma g(Wfxt + Ufht-1 + bf)$$

$it = \sigma g(Wixt + Uiht{-}1 + bi)$

$ot = \sigma g(Woxt + Uoht{-}1 + bo)$

$ct = ft \circ ct{-}1 + it \circ \sigma c(Wcxt + Ucht{-}1 + bc)$ $ht = ot \circ \sigma h(ct)$

Equation breakdown:

Forget Gate (ft)

$ft = \sigma g(Wfxt + Ufht{-}1 + bf)$

Where;

W f: Weight matrix for the input xt

Uf: Weight matrix for the previous hidden state $ht{-}1$

bf: Bias term.

σg: Sigmoid activation function, which outputs values between 0 and 1.

Input Gate (it)

$it = \sigma g(Wixt + Uiht{-}1 + bi)$

Where;

Wi: Weight matrix for the input xt

Ui: Weight matrix for the previous hidden state $ht{-}1$

bi: Bias term.

σg: Sigmoid activation function

Output Gate (ot)

$ot = \sigma g(Woxt + Uoht{-}1 + bo)$

Where;

W o: Weight matrix for the input xt

Uo: Weight matrix for the previous hidden state $ht{-}1$

bo: Bias term.

σg : Sigmoid activation function.

Cell State (ct)

$$ct = ft \circ ct{-}1 + it \circ \sigma c(Wcxt + Ucht{-}1 + bc)$$

Where;

f t: Forget gate output.

$ct{-}1$: Previous cell state.

it: Input gate output.

Wc: Weight matrix for the input xt

Uc: Weight matrix for the previous hidden state $ht{-}1$.

bc: Bias term.

σc : Activation function (usually tanh).

∘: Element-wise multiplication

Hidden State (ht)

$$ht = ot \circ \sigma h(ct)$$

where;

ot: Output gate output.

σh: Activation function (usually tanh).

∘: Element-wise multiplication.

3.3.2 ConvLSTM Model

The ConvLSTM model features a dual-branch architecture, consisting of an encoding branch and a forecasting branch, as illustrated in Figure 5. This design allows the model to process and predict sequential data with spatial and temporal dimensions effectively.

Encoding Branch: The purpose of the encoding branch is to compress the input sequence into a compact representation. This is achieved by passing the input through multiple ConvLSTM cells, which progressively encode the input sequence's spatio-temporal features into a hidden state tensor. Each cell in this branch processes the input while considering both spatial relationships (via convolution operations) and temporal dependencies (through LSTM mechanisms). The output of the last cell in this branch, including its state, encapsulates the entire input sequence's information (Shiet al., 2015).

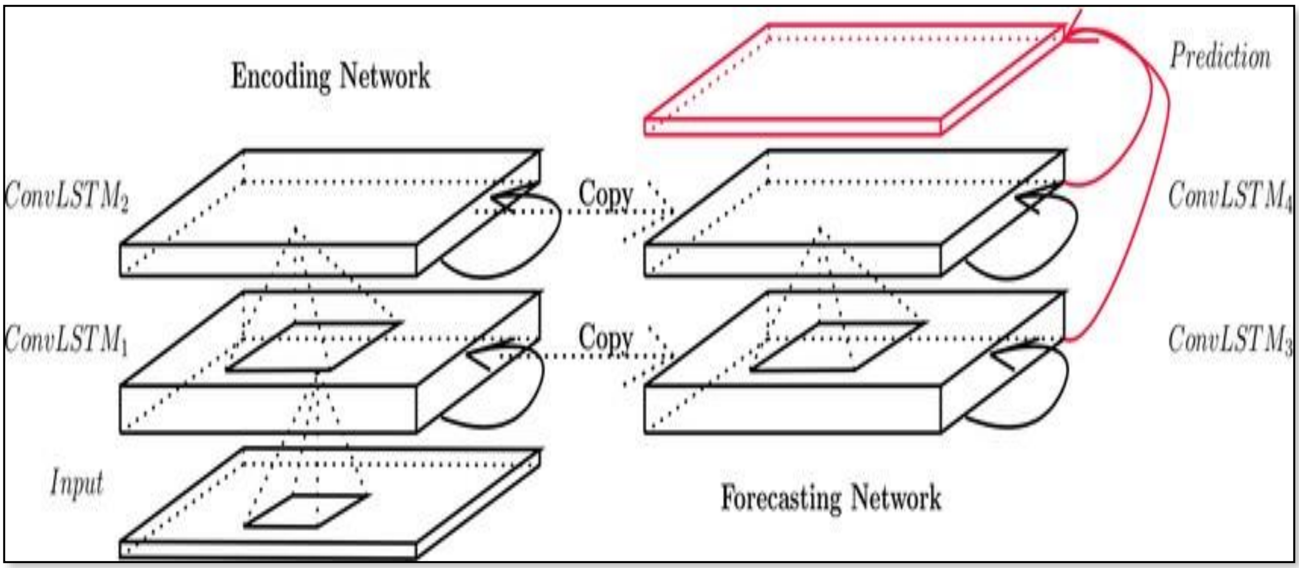

Figure 6: Inner Visualization of ConvLSTM structure encoding-Images to ConvLSTM network for Frame Prediction

Forecasting Branch: The forecasting branch assumes control after the encoding branch completes its task. It inherits the final cell output and its state from the encoding branch, utilizing them as the initial state and cell output for its operations. The primary objective of the forecasting branch is to reverse the encoding process, reconstructing the encoded hidden state tensor into a sequence. However, its focus shifts towards predicting future states based on the acquired spatio-temporal dynamics. Employing a series of ConvLSTM cells, the forecasting branch progressively expands the condensed information, advancing step by step to generate the final prediction sequence (Shi et al., 2015)..

The integration of multiple ConvLSTM cells across both branches enables the model to extract and leverage complex spatio-temporal features from the data, facilitating accurate predictions. This structured approach, dividing the task into encoding and forecasting phases, allows for efficient handling of sequences with intricate spatial and temporal patterns, making ConvLSTM models particularly suited for tasks like weather forecasting, video frame prediction, and any other application requiring nuanced understanding of space and time (Cho et al., 2014 & Karpathy et al., 2015).

The ConvLSTM model represents an advanced machine learning framework capable of processing video data, which encompasses both spatial and temporal information. This model modifies the conventional (LSTM) network, a type of time series Recurrent Neural Network, to better handle multidimensional data. Unlike the traditional Fully Connected Long Short-Term (FC-LSTM) networks, where input is limited to one-dimensional time-series vectors, ConvLSTMs adapt to manage image sequences by integrating convolution operations within its gates and states.

This approach allows for direct processing of spatial features alongside temporal dynamics. An alternate strategy seen in other studies involves sequentially combining a Convolutional Neural Network (CNN) with an LSTM network to form a CNN-LSTM architecture (9). This modular design first extracts spatial features via CNN before analyzing temporal sequences with LSTM, showcasing a different method of incorporating convolution into LSTM models for handling time-series image data.

ConvLSTM (Shi et al., 2015, Mendel et al., 2016, Chong et al., 2017) is a hybrid model designed to simultaneously extract both spatial and temporal features from spatio-temporal data. It achieves this by modifying the standard LSTM structure to incorporate convolution operations in the transitions from input to hidden states and from hidden-to-hidden states. This modification enables the ConvLSTM to carry forward spatial information across time through each state of the ConvLSTM network. By applying convolutional operations, the model efficiently maintains the spatial relationships within the data while also capturing the temporal dynamics, allowing for a comprehensive analysis of spatio-temporal datasets.

3..4 Input Data Labelling

To prepare our ConvLSTM input, we opted to concentrate solely on the blue cloud-like formations representing Nitrogen Dioxide. Initially, we imported our dataset into a JPEG format using the OpenCV Python package. Given the substantial size of the dataset, we divided it into two batches, each with distinct lower resolutions. One batch consisted of a 400px-by-400px JPEG image dataset comprising 169 images, while the other comprised a 40px-by-40px JPEG imagedataset, also containing 169 images. For both lower resolution sets, we applied a mask to isolate the (0,60,60) to (225,255,255) RGB color range, encompassing various light blue shades indicative of Nitrogen Dioxide. All other colors outside this range were set to the RGB color scheme (0,0,0) or black as shown in figure 3.

We generated four variations of the original image dataset: two datasets with a resolution of 40px by 40px and two with a resolution of 400px by 400px, each subjected to masking. Among these datasets, two underwent binarization, converting light blue pixels to 1 and black pixels to 0. Consequently, these two datasets comprised binary arrays. The remaining two datasets retained the masked color scheme, preserving the light blue and black color differentials. As a result, we possessed two sets of datasets containing color images in RGB format and two sets containing binary arrays. The utilization of color image datasets allowed for the application of a Rectified Linear Unit (ReLU) activation function in the output layer, while the binary array datasets facilitated the use of a sigmoid activation function. This approach broadened our scope for analysis and model training, enabling a more comprehensive exploration of the dataset's characteristics and features.

A pivotal determinant of the effectiveness and practicality of a time series model hinges on how we delineate the training and testing datasets from an uninterrupted flow of time-indexed data. In our methodology, each training set input was accompanied by the frame that immediately succeeds the current frame, fostering the model's ability to discern underlying patterns with the aim of generating precise predictions of future frames, specifically extending the forecast horizon to 13 days.

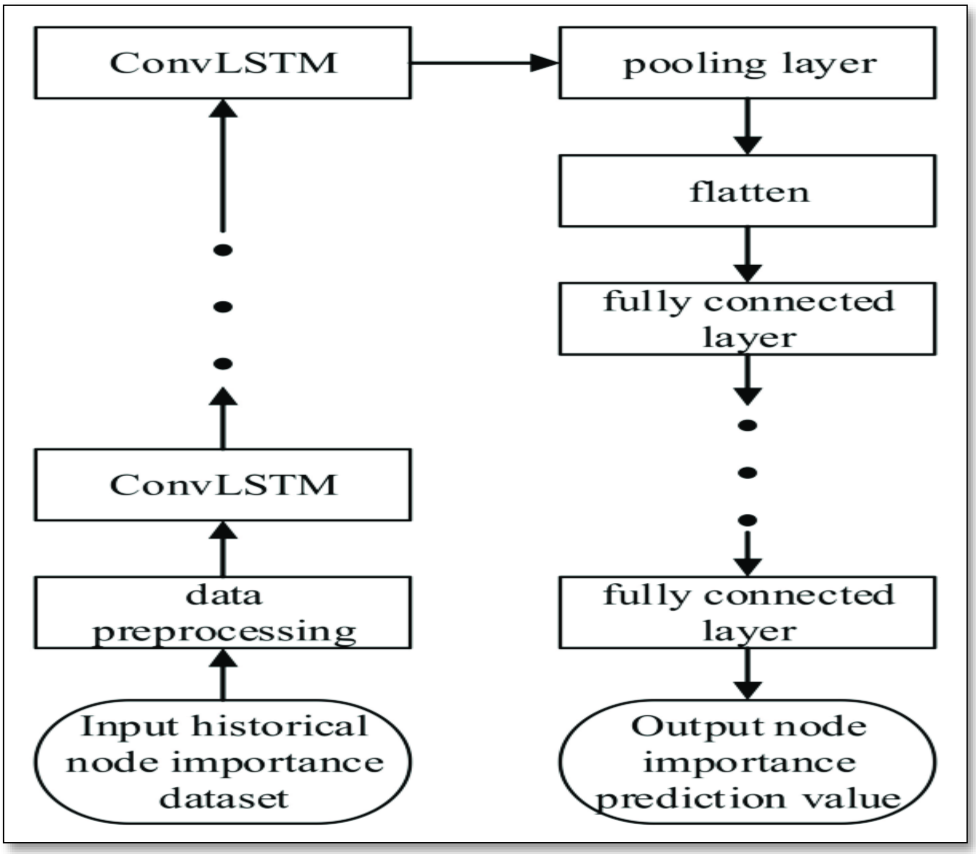

Figure 7: Operational flow of ConvLSTM

The flow diagram (Figure 7) illustrates the process of predicting node importance using a ConvLSTM-based approach. The steps involved are as follows:

- Input Historical Node Importance Dataset: The process begins with the input of a historical node importance dataset. This dataset contains historical data that is crucial for training the model.
- Data Preprocessing: The input data undergoes preprocessing to ensure it is in a suitable format for the model. This step may involve normalization, handling missing values, and other data cleaning tasks.
- ConvLSTM Layers: The preprocessed data is then fed into a series of ConvLSTM (Convolutional Long Short-Term Memory) layers. These layers are designed to capture both spatial and temporal dependencies in the data. The ConvLSTM layers process the data iteratively, allowing the model to learn complex patterns.
- Pooling Layer: After the ConvLSTM layers, the data is passed through a pooling layer. The pooling layer reduces the dimensionality of the data,

making the subsequent computations more efficient and helping to prevent overfitting.

- Flatten Layer: The output from the pooling layer is then flattened into a one-dimensional array. This step prepares the data for the fully connected layers by converting the multi-dimensional output into a format suitable for dense layer processing.

- Fully Connected Layers: The flattened data is passed through a series of fully connected (dense) layers. These layers perform the final stages of computation, allowing the model to learn intricate relationships in the data.

- Output Node Importance Prediction Value: The final output from the fully connected layers is the predicted node importance value. This value represents the model's prediction of the importance of a given node based on the input historical data.

This structured approach leverages the strengths of ConvLSTM networks in handling spatio-temporal data, providing a robust method for predicting node importance. We had created four resampling's of the original image dataset: two masked 40px by 40px datasets and two masked 400px by 400px datasets. For two of the datasets with a mask, one 40px by 40px and one 400px by 400px, we binarized the data, so that all light blue pixels became 1, and all black pixels became 0. Thus, two of our datasets were binary arrays. For the remaining two image datasets, we kept the masked light blue and black color scheme intact. In this way, we had two datasets of masked color images (RGB) and two datasets of bit arrays. The color image datasets allowed us to use a ReLU activation function in the output layer, while the binary array datasets allowed us to use a sigmoid activation function.

After resampling the data into various groups based on binary image, masked image, or image resolution criterion, we began parsing our 225 image datasets into readable formats for our ConvLSTM models. In the current form, our image dataset was in the form of (225,400,400,3) for 400px-by-400px masked images, (225,40,40,3) for 40px by 40px masked images, (225,400,400,1) for 400px by 400px binary images, and (225,40,40,1) for 40px by 40px binary images. We then took all our datasets and batched every five-image frame as one sample.

3.5 Data processing

To prepare the raw satellite images for processing with the ConvLSTM model, it was necessary to convert the data into a five-dimensional (5D) tensor format. This conversion process is crucial for aligning the data with the input requirements of ConvLSTM, which is designed to handle inputs with spatial and temporal dimensions.

The dataset comprised 169 high-resolution JPEG images, representing 845 days of observations. Given that the Sentinel-2 satellite has a revisiting time of 5 days, dividing the total number of observation days (845) by the revisiting time yields approximately 169. This calculation suggests that, over the observed period, Sentinel-2 could have potentially captured the same area about 137 times, allowing for comprehensive coverage and monitoring.

To accommodate the model's requirements and enhance processing efficiency, the original Sentinel-2 images, which have a resolution of about 7500 x 7700 pixels, were resized to a more manageable resolution of 400 x 400 pixels. This resizing step is crucial for standardizing the input data and ensuring the model can efficiently learn from the images. Additionally, the resized dataset was further divided into different subsets to facilitate training and validation of the ConvLSTM model, enabling a thorough assessment of its performance on spatiotemporal satellite imagery data.

For processing the satellite images with the ConvLSTM model, a sliding window approach was employed to structure the input and output data for the model. This method involves selecting sets of consecutive frames from the dataset to create distinct input-output pairs for training and validation.

Input Formation: Each input instance is composed of 5 consecutiveframes. These frames are grouped together to form a single input window, providing the model with both spatial and temporal information about the environment at a given time.

Output Formation: For each input window, the subsequent frame— immediately following the last frame of the input window—is used as the output frame. This output frame serves as the target for the model to predict, representing the immediate future state based on the information provided by the input frames.

Sliding Window Mechanism: After forming an input-output pair, the window is slid by a factor of one frame forward. This means the next input window will include the next 4 consecutive frames from the dataset plus the frame that was previously used as an output. This process is repeated, moving the window across the entire dataset, to generate multiple input- output pairs until all images are covered.

This approach ensures that each input example, consisting of 5 continuous frames, is directly linked to a subsequent single frame as the output example. The output frame acts as the "ground truth" or actual outcome against which the model's predictions are compared. This structured method of creating input-output pairs allows the ConvLSTM model to learn how to predict the next frame in a sequence based on the spatial and temporal information contained in the preceding frames, facilitating the model's training on the spatio-temporal dynamics captured in the satellite imagery.

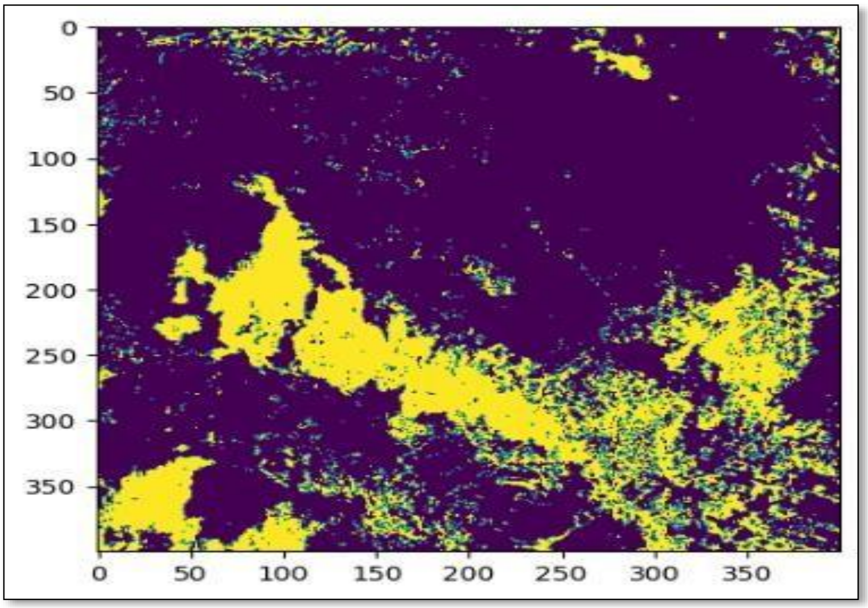

Figure 8: Processed satellite image (NO2)

3.5 Input Data Labelling for processing imagery

An important determinant of the success and practicality of a time series model is based on the selection of the training and testing sets from a continuous stream of time-indexed data. In our model, the label for a training set input is the frame that directly follows the current frame. In this way, the model is picking up on the pattern to base a current input image on the goal of accurately predicting a future image which in this case means predicting the next 10 days in the future.

In our model, we use five input frames as a data sample to predict five future frames. In fact, since each frame was separated by 2 days, we use 10 days of previous satellite imagery data to predict 10 days in the future of satellite imagery data. For the input labels, we first shifted over all data by five frames, with the first 5 frames corresponding to the 6th through 10th frames and the last 5 frames corresponding to the 226th through 230th frame. The 226th through 230th frames were filled with an averaging of the 30 days prior. Thus, we encoded sequential samples in training/testing inputs indexed by time and training/testing labels indexed by time, but did not overlap the training/testing inputs and the training/testing labels. We were able to perform a continuous prediction from the model's learned patterns.

3.6 Prediction using basic ConvLSTM Model

The architecture depicted in Figure 9 processes a sequence of sentinel 2 (No2) consecutive frames through a multi-layer ConvLSTM network. Each step of the architecture is designed to capture and enhance different spatial and temporal features from the input sequence, utilizing varying kernel sizes and numbers of filters across layers.

First ConvLSTM Layer: The initial layer receives the sequence of N frames and transforms them into a tensor. This layer applies a convolution operation using a kernel size of 5x5 and 64 filters. The convolution operation in this context is tailored to extract spatial features while also capturing the temporal dynamics across the sequence. The output of this layer is a feature-rich representation of the input sequence.

Subsequent ConvLSTM Layers: The output from the first layer is then passed to the next ConvLSTM layer. This layer further processes the data, potentially using a different kernel size (e.g., 3x3) and a reduced number of filters (e.g., 32), to refine the feature representation. This process is repeated in another layer, maintaining the kernel size of 3x3 but further reducing the number of filters to 16. Each layer's purpose is to progressively enhance the model's understanding of the input data's spatial and temporal aspects.

Final Convolution Layer: After the last ConvLSTM layer, a final convolution layer is applied to reconstruct the predicted image. This layer uses a kernel size of 3x3, three filters (corresponding to the RGB channels for color images, if applicable), and a stride of N. The choice of stride, kernel size, and filters in this layer is crucial for accurately reconstructing the predicted output frame from the processed feature maps.

Throughout the architecture, it is essential to meticulously monitor the dimensions of the inputs and outputs at each layer, ensuring compatibility and effective feature transformation. The specified kernel sizes, numbers of filters, and strides are selected to optimize the network's performance in extracting and processing spatio-temporal features from the input sequence, leading to accurate image prediction. This systematic approach allows the ConvLSTM network to learn and predict complex spatial and temporal dynamics present in sequences of high-resolution images.

3.7 Motivation for model choice

A strength of ConvLSTM lies in its ability to analyze both the spatial and temporal features within time series imagery, rendering it a powerful tool for forecasting future frames. This unique attribute allows ConvLSTM to excel in extracting information from sequences of images over time. Consequently, ConvLSTMs stand out as superior models for deriving insights from time series images, offering their use as predictive tools to anticipate upcoming frames. By integrating both spatial detail and temporal progression, ConvLSTMs provide a robust framework for forecasting, making them particularly useful in applications requiring precise future state predictions based on past and present imagery data.

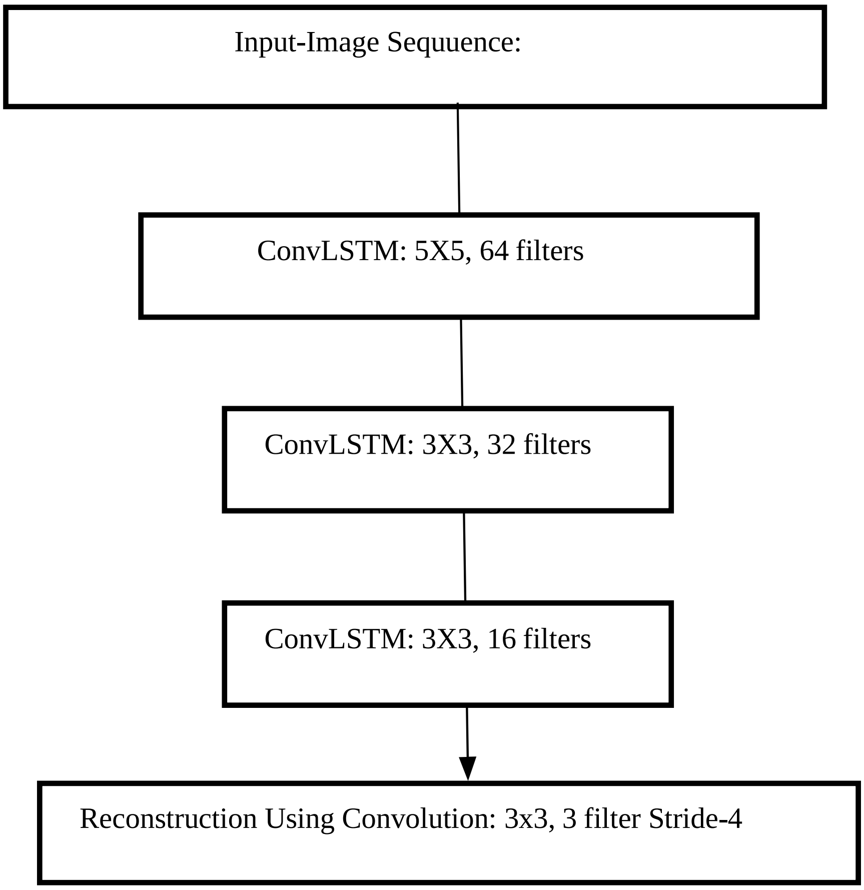

Figure 9: ConvLSTM Model operational inputs

3.8 Model operation from Python code

```python
# Define your ConvLSTM2D model
seq = Sequential()
seq.add(ConvLSTM2D(filters=1, kernel_size=(2, 2),
                   input_shape=(5, 40, 40, 1),
                   padding='same', return_sequences=True))
seq.add(BatchNormalization())

seq.add(ConvLSTM2D(filters=1, kernel_size=(3, 3),
                   padding='same', return_sequences=True))
seq.add(BatchNormalization())

seq.add(Conv3D(filters=1, kernel_size=(3, 3, 3),
               activation='sigmoid',
               padding='same', data_format='channels_last'))

# Compile the model with the desired optimizer, loss function, and metrics
seq.compile(loss='mean_absolute_error', optimizer='adam')

# Now you can fit your model to the data
seq.fit(X_train, y_train, batch_size=10, epochs=65, validation_split=0.1)

```

Figure 10: Model operational input codes (Python)

After a sequential model is initialized, Two ConvLSTM2D layers are added sequentially. ConvLSTM2D layers combine convolutional operations with LSTM (Long Short-Term Memory) units, making them capable of capturing spatial and temporal dependencies in the data.

The first ConvLSTM2D layer has 1 filter and a kernel size of (2, 2). It accepts input data of shape (5, 40, 40, 1), which could represent a sequence of 5 frames, each of size 40x40 pixels with a single channel (e.g., grayscale images). The 'same' padding ensures that the output has the same spatial dimensions as the input, and return_sequences=True means that the layer will return the full sequence of outputs for the next layer to process.

Batch-Normalization is applied after each ConvLSTM2D layer to normalize the outputs, which helps stabilize and accelerate the training process.

3.8.1 Adding Conv3D Layer:

A Conv3D layer is added next. Conv3D layers perform 3D convolutions, which are suitable for processing volumetric data (e.g., 3D images or sequences of 2D images).

This layer has 1 filter and a kernel size of (3, 3, 3). The 'sigmoid' activation function is applied to the output, which squashes the output to a range between 0 and 1. The 'same' padding ensures that the output maintains the same spatial dimensions as the input.

3.8.2 Model Compilation:

The model is compiled with the 'mean_absolute_error' (MAE) loss function, which measures the average magnitude of errors between predicted and actual values. The 'adam' optimizer is used for training the model. Adam is an adaptive learning rate optimization algorithm that has been widely adopted due to its efficiency and performance.

3.8.3 Model Training

The model is trained on the dataset X_train with corresponding labels y_train. The batch_size parameter is set to 10, meaning the model will process 10 samples at a time before updating the weights. The epochs parameter is set to 65, indicating that the model will go through the entire training dataset 65 times. A validation_split of 0.1 means that 10% of the training data will be set aside for validation, allowing the model's performance to be evaluated on unseen data during training.

The code depicted in figure 10 defines a neural network model using ConvLSTM2D layers to handle spatiotemporal data, compiles the model with appropriate loss and optimization functions, and trains the model on a given dataset. This model is designed to learn and make predictions based on input data that has both spatial and temporal characteristics.

3.8.4 Model Implementation

In this ConvLSTM model setup, the input layer receives a sequence of 5 consecutive images, each with dimensions of 304x40x40x1. These images undergo a series of transformations through ConvLSTM layers, each designed to extract and refine spatiotemporal features for future frame prediction.

First ConvLSTM Layer: This layer focuses on extracting spatial features via convolution, using a filter size of 5x5 and 64 filters. Concurrently, it captures temporal dynamics through LSTM mechanisms. The output remains a sequence of 5 images, each retaining the input's dimensions (5x40x40x1), indicating that spatial details are emphasized while temporal relationships are maintained. Batch normalization is applied to this output to standardize and stabilize the learning process.

Second ConvLSTM Layer: The processed output from the first layer is forwarded to the next ConvLSTM layer, which further extracts spatiotemporal features. This layer employs a smaller 3x3 kernel size with 32 filters, adjusting the feature representation (noted as output dimension (140) which seems to be a typographical error and might represent a sequence or a feature map size). Batch normalization is again applied to ensure smooth training and effective feature representation.

Third ConvLSTM Layer: Continuing the feature refinement, this layer utilizes a 3x3 kernel size with 16 filters to further distill the spatiotemporal information from the data (noted as output dimension (148), which might also be a typographical or formatting error related to feature maps or sequence representation).

Final Convolution Layer: The culmination of the process involves a convolution operation on the last layer's output using a 3x3 filter size, 3 filters, and a stride of 5. This step is designed to reconstruct a single forecasted image with dimensions of 5x40x40x1, as illustrated in Figure 10.

The model's forecasted image is then compared against the 6th image in the sequence, which serves as the ground truth for model evaluation. This methodology showcases the layered approach of ConvLSTM in handling complex spatiotemporal data, progressively focusing the model’s learning on predicting future states based on the accumulated spatial and temporal insights. Figure 11: Snapshot of model where the input layer consists of five consecutive frames, each with dimensions of 40x40x1. These frames pass through subsequent layers, ultimately producing a single predicted future frame.

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv_lstm2d_2 (ConvLSTM2D)  (None, 5, 40, 40, 1)      36

 batch_normalization_2 (Bat  (None, 5, 40, 40, 1)      4
 chNormalization)

 conv_lstm2d_3 (ConvLSTM2D)  (None, 5, 40, 40, 1)      76

 batch_normalization_3 (Bat  (None, 5, 40, 40, 1)      4
 chNormalization)

 conv3d_1 (Conv3D)           (None, 5, 40, 40, 1)      28

=================================================================
Total params: 148 (592.00 Byte)
Trainable params: 144 (576.00 Byte)
Non-trainable params: 4 (16.00 Byte)
_________________________________________________________________
```

Figure 11: Sequential model operation

CHAPTER 4

Results and Analysis

4.1 Result

4.1.1 ConvLSTM2D

The initial row in Table 1 indicates that each input example consists of 5 consecutive images with a resolution of 40x40x1. For training and validating the model, only Sentinel-2 data was utilized, resulting in 163 training examples and 7 validation examples. These examples were processed through a 3-layer basic ConvLSTM model with a batch size of 5, trained over 25 epochs, yielding a training error of 0.2209 and a validation error of 0.1714. The outcomes generated by this approach are illustrated in Figure 11.

Similarly , the second row of Table 1 has a window size of 5 sequential images of purely Sentinel-2 data having 164 training examples and 11 validation examples with 2 layers of basic ConvLSTM gives training error loss: 0.1758 and a validation error val_loss: 0.1144. Table one shows the results of the model after running 3 groups oof epochs. The model performance is proportional to the number of determined epochs. Figure 12 shows the physical presentation of pollution prediction against ground data. Frame 1 is for 25 epochs and frame 5 is for 100 epochs.

Input	Data Type	Training examples	Validation examples	layers	Batch Size	Epochs	Training Err Val_loss
169,40,40,1	Sentinel-2	110	7	3	5	25	loss: 0.2209 - val_loss: 0.17
169,40,40,1	Sentinel-2	110	7	3	5	50	loss: 0.1758 - val_loss: 0.11
169,40,40,1	Sentinel-2	110	7	3	5	100	loss: 0.1704 - val_loss: 0.11

Table 1: The table displays various combinations of hyperparameters alongside their respective training and validation losses

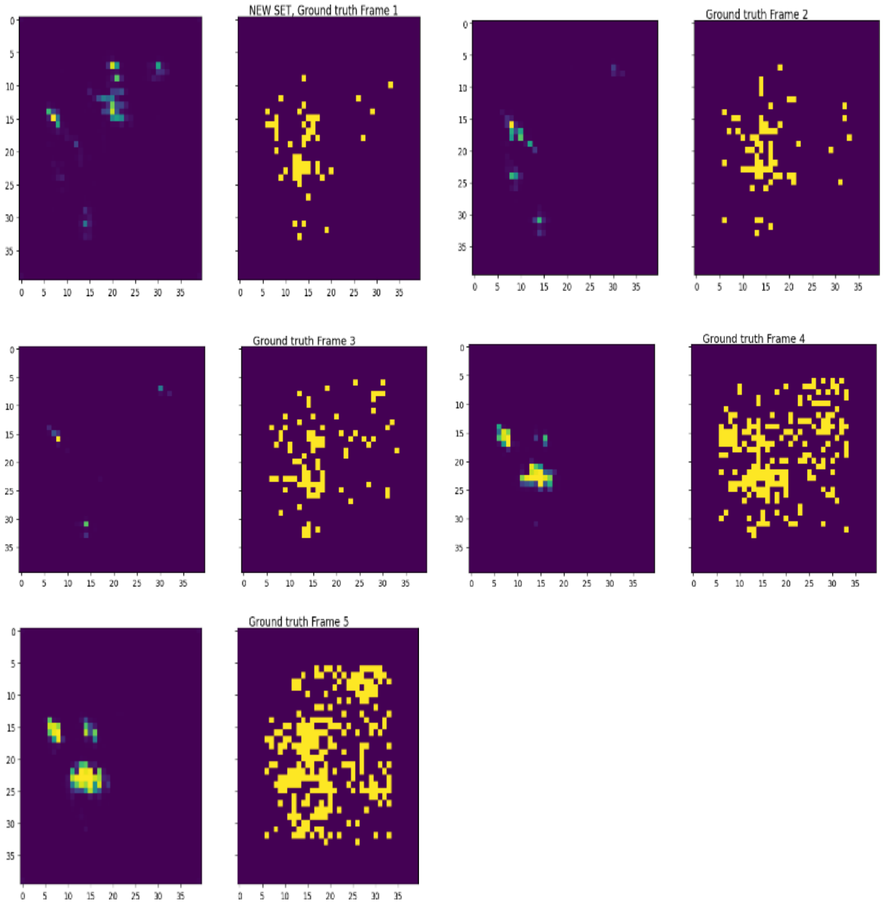

Figure 12: The output represents the performance outcomes of a basic ConvLSTM model with three layers when fed with five consecutive frames representing 13 days as input.

The two figures (12 & 13) corresponds to a time point approximately 13-17 days apart. In this context, the model forecasted events occurring around 1 month into the future, specifically targeting the time frame around January 2024.

In figure 12 the result of predicted and ground truth frame are shown. This was after 50 epochs were implemented. The model can not clearly mimic the ground truth data and it led to misclassification of the predicted frames. The same issue persists in figure 13, which was ran after 100 epochs, where in the last two frames (4 & 5), the frames were misclassified.

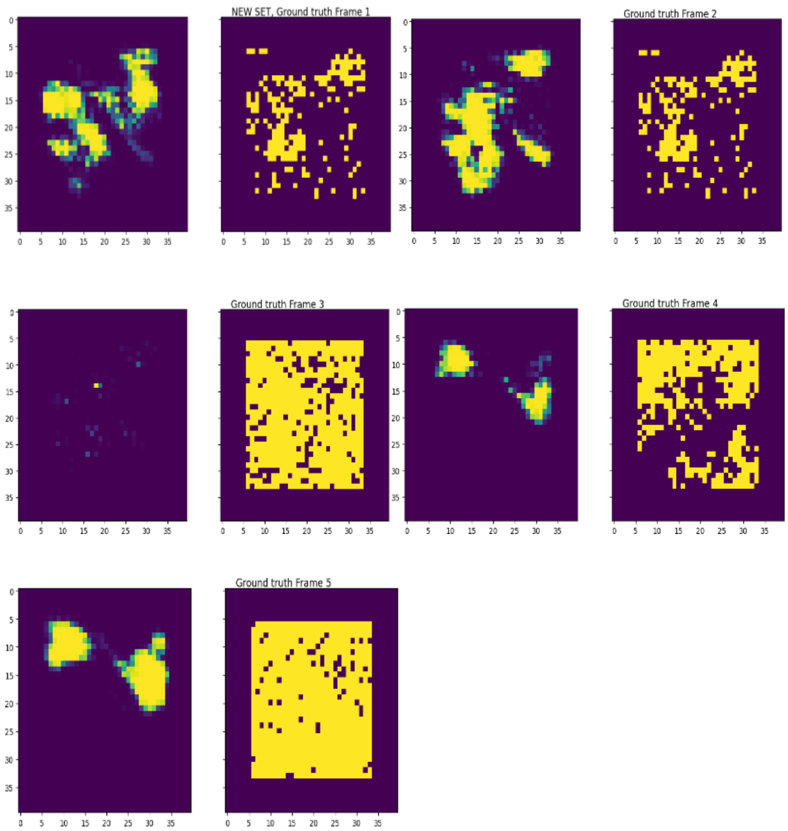

Figure 13: The output signifies the predictions made by the ConvLSTM model for 10 future frames

4.1.2 Error analysis

Given that our model produces a sequence of images depicting atmospheric nitrogen dioxide levels over Eastern Gauteng and western Mpumalanga, conventional error assessment methods do not accurately capture the model's precision. Calculating the root mean square error (RMSE) on a pixel-by-pixel basis between actual and forecasted images fails to adequately represent the congruence in the nitrogen dioxide structures depicted across the images.

Opting for the Structural Similarity Index Measure (SSIM) as an alternative to traditional error metrics like the root mean square error (RMSE) marks a strategic approach to assessing the precision of your model in forecasting atmospheric nitrogen dioxide levels. SSIM, widely recognized in computer vision and image analysis fields (Woo et al., 2014), offers a nuanced evaluation of image quality and similarity by comparing the structural integrity of two images rather than just their pixel-by-pixel brightness values.

Unlike RMSE, which calculates the average magnitude of errors between corresponding pixels in the actual and predicted images, SSIM delves deeper into the visual composition of the images. It assesses changes in texture, contrast, and luminance by evaluating the pixel averages and standard deviations within localized patches of the images. By doing so, SSIM effectively captures the perceptual differences between the forecasted and actual images, including the spatial distribution and intensity patterns of nitrogen dioxide (Zhou et al., 2004).

This method is particularly beneficial for environmental modelling, where the spatial arrangement and continuity of atmospheric phenomena are more indicative of model accuracy than isolated pixel value comparisons. By employing SSIM, you can gain insights into how well the model preserves the structural and textural integrity of nitrogen dioxide distributions, offering a more comprehensive understanding of its predictive capabilities.

The SSIM metric is defined as:

$$\mathrm{SSIM}(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}$$

Equation 5:SSIM

Where:

- x and y are the two images being compared.
- μ_x and μ_y are the average values (mean) of x and y.
- σ_x^2 and σ_y^2 are the variances of x and y.
- σ_{xy} is the covariance between x and y.
- C_1 and C_2 are small constants used to stabilize the division with weak denominators.

In this context, I represents the ground truth images, I denotes the predicted images, p refers to the pixels of the ground truth images, and $p\hat{}$ signifies the pixels of the predicted images. μ stands for the average pixel value of each image, σ indicates the standard deviation of pixel values in each image, and $c1$ and $c2$ are constants related to the relative noise present in the images (Lin et al., 2011).

Table 2 displays the SSIM error values between the predicted outputs and the ground truth for the initial five frames of the testing set, covering approximately a 10-day period into the future.

ConvLSTM Model over sample 2		
	SSIM	MSE
Frame 1	0.34	0.11
Frame 2	0.40	0.19
Frame 3	0.07	0.28
Frame 4	0.30	0.17
Frame 5	0.22	0.08
Sample Difference (5Frames)		
	0.19	0.17

Table 2: Error analysis :The SSIM & MSE values for the first sample (first 5 frames) represent the percentage of structural similarity between the predictions of Nitrogen Dioxide levels in Secunda for ten days into the future

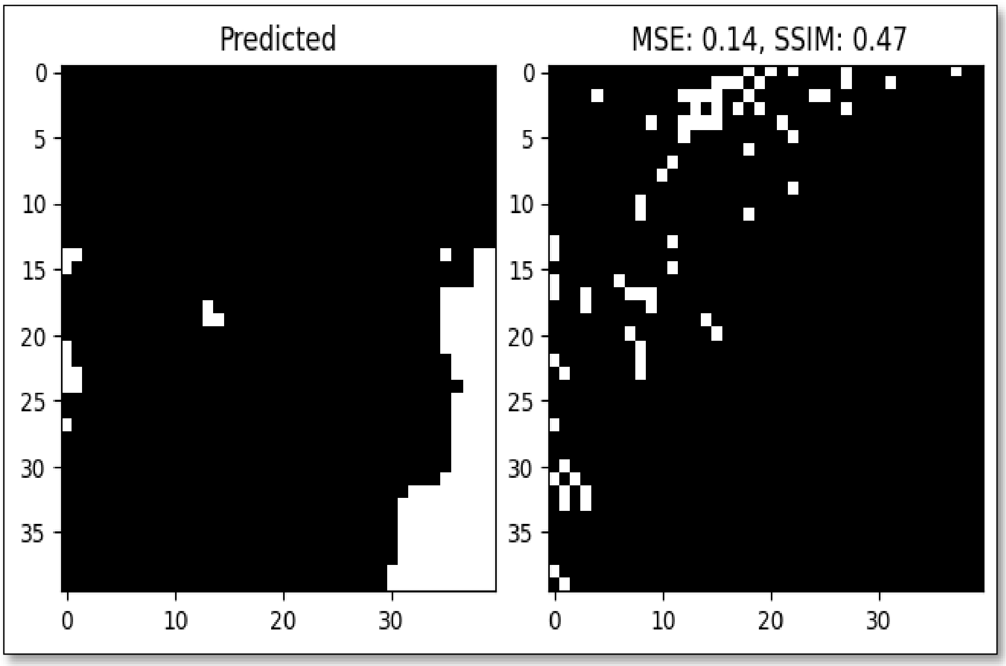

Figure 14: Predictive error analysis over FIRST Frame

The SSIM error metric indicates a 47% structural similarity between the predicted and ground truth images of Nitrogen Dioxide air pollution in Secunda. This prediction is based on binary input data from the previous 10 days.

The ConvLSTM model demonstrates variable performance across the sequence. While MSE values fluctuate, indicating varying degrees of pixel-level accuracy, SSIM values consistently indicate moderate to significant differences in structural content compared to the reference frames. These metrics suggest that the model may benefit from further training or adjustment to better capture the structural features of NO2 distribution patterns.

The accuracy of predicting future frames decreases notably for higher values, primarily because the ground truth exhibits less correlation with the nitrogen dioxide levels of the preceding 10 days compared to the initial frames of the ground truth sequence

The model's Structural Similarity Index (SSIM) exhibits instability, fluctuating based on the frame it predicts. This poses a significant challenge within our research. Some frames demonstrate high SSIM values, while others show lower scores. This fluctuation may stem from the quality of the data utilized, as we did not selectively curate it, encompassing both cloudy and clear images.

4.1.3 ConvLSTM model with Meteorological data

We propose to continuously predict the whole station x feature map in the next 1 to 48 hours with a non-stationary convolutional LSTM (), which replaces the fully connected network in the classical LSTM with detrending and convolutional operations. The LSTM model we used was quite common. We aggregated air pollution values by hour and considered each hourly value as a point in the time series sequence.

#Snap code

Seq2seq

```python

seq = Sequential()
seq.add(ConvLSTM2D(filters=1, kernel_size=(2, 2),
                   input_shape=(5, 35, 1, 1),  # Adjusted to match spatial data
                   padding='same', return_sequences=True))
seq.add(BatchNormalization())

# ... rest of the model

seq.add(ConvLSTM2D(filters=1, kernel_size=(2, 2),
                   padding='same', return_sequences=True))
seq.add(BatchNormalization())

seq.add(Conv3D(filters=1, kernel_size=(3, 3, 3),
               activation='sigmoid',
               padding='same', data_format='channels_last'))

# Compile the model
seq.compile(loss='mean_absolute_error', optimizer='sgd')

```

Figure 15: Model code (Python)

4.1.3.1 General Overview:

A Sequential model is initialized. The Sequential model allows for the linear stacking of layers, which is suitable for building a straightforward feed-forward neural network

First ConvLSTM2D Layer:

A ConvLSTM2D layer is added with 1 filter and a kernel size of (2, 2). The input_shape is set to (5, 35, 35, 1), indicating a sequence of 5 frames, each with dimensions of 35x35 pixels and 1 channel (e.g., grayscale images). padding='same' ensures that the output has the same spatial dimensions as the input. return_sequences=True means that the layer will return the full sequence of outputs, which is necessary for the next ConvLSTM2D layer to process. BatchNormalization is applied to stabilize and accelerate the training process by normalizing the output of the ConvLSTM2D layer.

Second ConvLSTM2D Layer:

A second ConvLSTM2D layer is added with the same configuration as the first, but without specifying input_shape since it inherits the input shape from the previous layer. BatchNormalization is applied again to normalize the outputs of this layer.

Conv3D Layer*:*A Conv3D layer is added with 1 filter and a kernel size of (3, 3, 3). This layer performs 3D convolutions, suitable for processing volumetric data or sequences of 2D images.

The activation function is set to 'sigmoid', which squashes the output values to a range between 0 and 1. padding='same' ensures that the output has the same spatial dimensions as the input. data_format='channels_last' indicates that the channels dimension is the last in the input data shape.

Model Compilation*:* The model is compiled with the 'mean_absolute_error' (MAE) loss function, which measures the average magnitude of errors between predicted and actual values.The 'sgd' (Stochastic Gradient Descent) optimizer is used for training the model. This code defines a sequential model using ConvLSTM2D layers to handle spatiotemporal data. The model is compiled with a specified loss function and optimizer, preparing it for training. The ConvLSTM2D layers capture spatial and temporal dependencies, while the Conv3D layer performs final 3D convolutions for the prediction task. Batch Normalization layers are included to stabilize and accelerate the training process.

Model summary

Input	Data Type	Training examples	validation examples	layers	Batch Size	Epochs	Training Error Loss Val_loss
5,35,1,1	Meteorological	104	5	3	5	25	loss: 0.1323 – val_loss: 0.1506
5,35,1,1	Meteorological	104	5	3	5	50	loss: 0.1287 – val_loss: 0.1502
5,35,1,1	Meteorological	104	5	3	5	100	loss: 0.1237 – val_loss: 0.1458

Table 3:The table displays various combinations of hyperparameters alongside

their respective training and validation losses

The model demonstrates a consistent decrease in both training and validation loss as the number of training epochs increases, which suggests that the model benefits from longer training periods without significant overfitting issues. It is noteworthy that the difference between training loss and validation loss narrows with more epochs, indicating improved generalization.

However, the validation loss does not show a dramatic decrease, which may suggest a point of diminishing returns where additional training provides increasingly smaller improvements. This also could imply that the model is nearing its capacity to learn from the current dataset and architecture configuration.

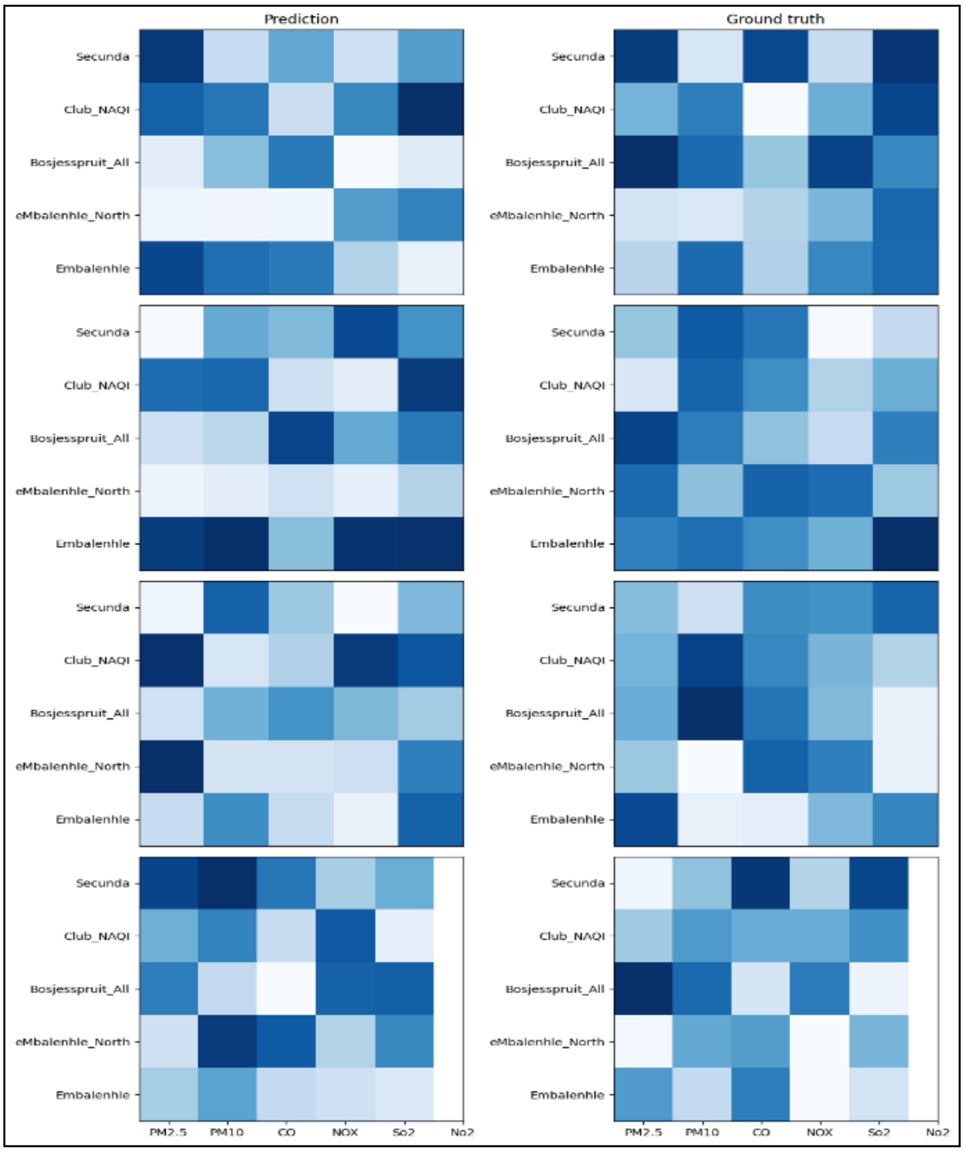

Figure 15: Predictive results for ConvLSTM with air monitoring data

The four heatmaps we generated represents data values generated by our predictive model in a matrix format, where colors of varying intensity correspond to the magnitude of the metric being observed. The choice for heatmap is its usefulness to visualizing complex data and identifying patterns, trends and correlations. In the above figure 16, the comparing predicted data against actual (ground truth) data for the next hour. The heatmap shows fluctuation in the predictive analysis with the use of color. The darker the blue color, the higher the concentration.

Chapter 5

5.1 Discussion

This study proposes a novel model to accurately predict air pollutant (No_2) within the refinery plant of Sasol in Secunda Mpumalanga. Most studies uses machine learning or deep learning models individually to predict air pollution using satellite data or air pollution data from monitoring weather stations. However this study leverages on the dual operation of recent sequential algorithms (ConvLstm) model which caters for both spatial and temporal analysis. This is a crucial aspect in air pollution prediction as air pollution is a spatial and temporal phenomenon. In this section , we discuss the insights and limitations of the model in context to this study.

First, the prediction of No_2, with sentinel 2 satellite imagery was poor as compared to those obtained in other relevant studies i.e (Pratyush Muthukumar., et al,2021). This can be visually seen in figures 12 and 13 of the study. The model performs well in the first and second frames by mimicking the patterns of ground truth data, however it loses the predictive performance in frame 3 to 5. The structural similarity index (SSIM) , which is a powerful metric in predictive analysis for evaluating how well a model preserves the structure and quality of the original data in its predictions was below +-.40 in average. An SSIM score of 0.40 reflects moderate structural similarity but with significant deviations from the original or expected structure, potentially indicating issues with the quality of the prediction or the effectiveness of the model.

Interestingly the MSE of the predictive analysis shows a very good performance, with an average of MSE of less than 0.25 in average. An MSE value of 0.25 indicates that the pixel values of the two images are closer, which generally implies higher similarity. An MSE of less than 0.30 is relatively low, suggesting that the pixel-wise differences between the images are minor.

It is worth noting that the Structural Similarity Index Measure (SSIM) was moderate, suggesting that, despite the numerical closeness indicated by the MSE, there were still perceptual and structural differences between the images. This finding underscores the importance of using both MSE and SSIM in tandem to fully assess the accuracy and quality of the predictions, as MSE alone may not capture all the relevant aspects of image fidelity.

This could have been caused by limited data used for analysis in the model as we *Second*, the predictive analysis of air pollution was expanded to other pollutants such as PM_{10}, $PM_{2.5,}$ $SO_{x,}$ CO and $No_{x.}$ The same model of sequential learning (ConvLstm) was implemented for 13 days prediction within the same area as the air monitoring station. The model loss and validation loss was faily low indicating good model performance within various epochs batches. The prediction was viewd in the form of a head map. Where some stations had high predictions than others i.e colors. Perhaps this good performance was because of the choice of data, meaning that in the model architecture performs well with numerical data than pictorial data.

This has fairly three major limitations. *First*, there was no consideration of the geo-coordinates in the data used for prediction. This leads to loss of spatial context as geo-coordinates provide essential spatial context that helps in understanding where specific pollution events are occurring. Without this information, the model might have missed critical spatial patterns and relationships, leading to less accurate predictions.

Secondly, it is common knowledge that deep learning models performs well with high quality and amount of data. In this study, we fell short of the data available and limited satellite data can significantly impact a CONVLSTM model's performance in air pollution prediction. Inadequate geographic and temporal coverage can lead to gaps in data, resulting in poor predictions, especially in regions with sparse information or during rapid pollution changes. This lack of comprehensive data must have hindered the model's ability to generalize across different scenarios, leading to biases and inaccuracies. Additionally, with insufficient data, the model might have suffered from underfitting, struggling to learn meaningful patterns and failing to capture fine-grained spatial and temporal variations in pollution levels.

Thirdly, since Sentinel-2 data does not include direct pollution metrics, any attempt to infer pollution levels from surface characteristics might lead to inaccuracies. The model could have misinterpreted surface features as proxies for pollution, potentially distorting predictions.

5.2 Conclusion

In this paper, we leverage the capabilities of Convolutional Long Short-Term Memory (ConvLSTM) deep learning model to analyse and forecast using high-resolution, multi-spectral time-series imagery from the Sentinel-2 satellite and meteorological data. ConvLSTMs uniquely combine the spatial feature extraction proficiency of convolutional neural networks (CNNs) with the sequential data processing strengths of (LSTM) networks. This fusion enables the effective capture of both spectral and temporal characteristics inherent in the satellite data, facilitating the accurate prediction of subsequent images.

Additionally, to complement the spatial and spectral insights obtained from Sentinel-2 imagery, we incorporated an LSTM model to analyse temporal meteorological data collected from air quality monitoring stations in Secunda. This model focuses on leveraging time-series data to enhance the temporal dimension of our analysis, offering a more holistic view of the environmental factors influencing air quality. Through a strategic combination of ConvLSTM for spatial and spectral analysis and LSTM for in-depth temporal analysis, our approach presents a comprehensive and efficient methodology for environmental monitoring and forecasting, with significant implications for urban planning and public health initiatives.

To comprehend both temporal and spatial patterns, we devised a Convolutional Long Short-Term Memory (ConvLSTM) model, a sophisticated machine learning technique tailored for datasets consisting of sequences of multi-dimensional frames. Our model effectively utilized 13 days of Nitrogen Dioxide air pollution data from the broader Los Angeles region to forecast Nitrogen Dioxide pollution levels across Secunda 13 days ahead.

In the analysis of nitrogen dioxide spatial distribution using Sentinel-2 data, the ConvLSTM model consistently exhibited a reduction in both training and validation loss with increasing epochs of training. This trend suggests that the model benefits from extended training periods without encountering significant overfitting problems, thus implying enhanced generalization capabilities.

However, even though the MSE decreased ±0.14, showing that the predictions of the model are closer to the actual values, suggesting better accuracy and performance of the model, the SSIM was lower, ±0.47, The instability of the Structural Similarity Index (SSIM) in the model varies depending on the predicted frame, presenting a notable challenge in our research. While certain frames show high SSIM values, others exhibit lower scores. This variability likely originates from the quality of the data, as it was not selectively curated, encompassing both cloudy and clear images.

With meteorological data, the four heatmaps generated as a result of ConvLSTM model for predictive vs ground truth data was more convincing that the spatial representation with sentinel-2 data. The training loss and validation loses proved that the model learned the data very effectively to avoid over-fitting and thus producing great results. It is evident that the ConvLSTM model performed well with air pollution data from monitoring data than with satellite data.

The application of ConvLSTM models in this context is especially pertinent for urban and regional authorities, as it provides a powerful tool for monitoring pollution cycles and trends over time. By harnessing the predictive insights generated by these models, cities and provincial governments can gain advanced warnings of potential environmental concerns, allowing for proactive measures to mitigate pollution before it reaches critical levels.

5.3 Future work

In our pursuit to improve the accuracy and precision of prediction of air pollution using deep learning models with satellite imagery, our hope is to develop a model that incorporates pollution-contributing factors as input variables and expanding pollutants to include Co2, PM2.5, SO2 and CH4. By integrating satellite data analyzed with ConvLSTM alongside meteorological data from monitoring stations, we envision creating a unified image presentation encompassing both spatial and temporal components.

Future iterations of these models will prioritize geographical coordinates and time, recognizing their influence on prediction accuracy. Additionally, we intend to implement a deep learning classification model to categorize predicted images into levels of pollution: high, low, and none. The future work will also expand area of interest and take into account two provinces of Gauteng and Mpumalanga.

References

1. Academy of Science of South Africa, et al. Air Pollution and Health – A Science- Policy Initiative. Annals of Global Health. 2019; 85(1): 140, 1–9. DOI: https://doi.org/10.5334/aogh.2656.

2. Aizebeokhai AP. Global Warming and Climate Change: Realities, Uncertainties, and Measures. International Journal of Physical Science. 2009;4(13):868-879.

3. Article 3.1 of the UNFCCC. Articles 1.7, 2.1 and 3.1 of the Kyoto Protocol. The Kyoto Protocol distinguishes between developed countries and those still developing. The developed countries are listed as Annexe 1 countries and have binding obligations in terms of the protocol; also see art 3.1- 3.2 of the UNFCCC.

4. Ballantyne AP, Andres R, Houghton R, Stocker BD, Wanninkhof R et al (2015) Audit of the global carbon budget: estimate errors and their impact on uptake uncertainty. Biogeosciences 12:2565–2584. https:// doi.org/10.5194/bg-12-2565-2015.

5. Barnwell, L., 2009. Industrial perspectives on the implementation of the Air Quality Act (AQA)(Act No. 39 of 2004) (Doctoral dissertation).

6. Bartholomew DJ. Time series analysis forecasting and control. J Oper Res Soc.1971;22(2):199–201. https://doi.org/ 10.1057/jors.1971.52.

7. Beckerman, B.; Jerrett, M.; Martin, R.V.; Van Donkelaar, A.; Ross, Z.; Burnett, R.T. Application of the deletion/substitution/addition algorithm to selecting land use regression models for interpolating air pollution measurements in California. Atmos. Environ. 2013, 77, 172–177.

8. Beelen, R.; Hoek, G.; Vienneau, D.; Eeftens, M.; Dimakopoulou, K.; Pedeli,

X.; Tsai, M.-Y.; Künzli, N.; Schikowski, T.; Marcon, A.; et al. Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe—The ESCAPE project. Atmos. Environ. 2013, 72, 10–23.

9. Bewket, W. (2012). Climate change perceptions and adaptive responsesof smallholder farmers in Central Highlands of Ethiopia. International Journal of Environmental Studies, 3(69):507–523.

10. Boesch, H.; Liu, Y.; Tamminen, J.; Yang, D.; Palmer, P.I.; Lindqvist, H.; Cai, Z.; Che, K.; Di Noia, A.; Feng, L.; et al. Monitoring (GHG'S) from Space. Remote Sens. 2021, 13, 2700. https:// doi.org/10.3390/rs13142700.

11. Botai, C.M., Botai, J.O. and Adeola, A.M., 2018. Spatial distribution of temporal precipitation contrasts in South Africa. South African Journal of Science, 114(7-8), pp.70-78.

12. Bougoudis, I., Demertzis, K. and Iliadis, L., 2016. HISYCOL a hybrid computational intelligence system for combined machine learning: the case of air pollution modeling in Athens. Neural Computing and Applications, 27, pp.1191-1206.

13. Bougoudis, I., Demertzis, K., Iliadis, L., Anezakis, V.D. and Papaleonidas, A., 2016. Semi-supervised hybrid modeling of atmospheric pollution in urban centers. In Engineering Applications of Neural Networks: 17th International Conference, EANN 2016, Aberdeen, UK, September 2-5, 2016, Proceedings 17 (pp. 51-63). Springer International Publishing.

14. Bovensmann, H., Buchwitz, M., Burrows, J.P., Reuter, M., Krings, T., Gerilowski, K., Schneising, O., Heymann, J., Tretner, A., Erzinger, J., 2010. A remote sensing technique for global monitoring of power plant CO2 emissions from space and related applications. Atmos. Measurement Tech. 3, 781–811. https://doi.org/ 10.5194/amt-3- 781- 2010.

15. Bruckner, T., Bashmakov, I., Mulugetta, Y., Chum, H., de la Vega Navarro, A., Edmonds, J., Faaij, A., Fungtammasan, B., Garg, A., Hertwich, E., Honnery, D., Infield, D., Kainuma, M., Khennas, S., Kim, S., Nimir, H., Riahi, K., & Strachan, N. (2022). Time series analysis of satellite data using ConvLSTM for spatio-temporal feature extraction and prediction. Indian Statistical Institute, Kolkata-700108, India.

16. Changeat, Q. and Yip, K.H., 2022. ESA-Ariel Data Challenge NeurIPS 2022: Introduction to exo-atmospheric studies and presentation of the Atmospheric Big Challenge (ABC) Database. arXiv preprint arXiv:2206.14633.

17. Chiou-Jye, H., & Ping-Huan, K. (2018). A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities. Sensors, 18(7), 2220.

18. Chiwewe, T. M., & Ditsela, J. (2019). Machine Learning Based Estimation of Ozone Using Spatio-Temporal Data from Air Quality Monitoring Stations. IBM Research South Africa, IBM Johannesburg, South Africa.

19. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. CoRR abs/1412.3555. Published 2014. Available from: https://paperswithcode.com/paper/empirical-evaluation-of-gated-recurrent.

20. Nassar, R., Mastrogiacomo, J.P., Bateman-Hemphill, W., McCracken, C., MacDonald, C.G., Hill, T., O'Dell, C.W., Kiel, M. and Crisp, D., 2021. Advances in quantifying power plant CO2 emissions with OCO-2. Remote Sensing of Environment, 264, p.112579.

21. Chong, Y.S. and Tay, Y.H., 2017. Abnormal event detection in videos using spatiotemporal autoencoder. In Advances in Neural Networks-ISNN 2017: 14th International Symposium, ISNN 2017, Sapporo, Hakodate, and Muroran, Hokkaido, Japan, June 21–26, 2017, Proceedings, Part II 14 (pp.

189-196). Springer International Publishing.

22. Climate change Bill. Section 2(b) of the Bill.

23. Climate Change Bill. Section 2(c) of the Bill

24. Climate change Bill. Section 2(c) of the Bill.

25. Coelho, S., Ferreira, J., Rodrigues, V. and Lopes, M., 2022. Source apportionment of air pollution in European urban areas: Lessons from the Clair City project. Journal of Environmental Management, 320, p.115899.

26. Coelho, S., Rafael, S., Lopes, D., Miranda, A.I. and Ferreira, J., 2021. How changing climate may influence air pollution control strategies for 2030?. Science of The Total Environment, 758, p.143911.

27. Darghouth, M., & Bouattour, A. (2008). Ticks and tick-borne diseases of livestock in North Africa, present state and potential changes in the context of global warming. In Livestock and Global Climate Change (pp. 112-118). Cambridge University Press.

28. Davidson, O., Tyani, L., Afrane-Okesse, Y., (2002), Climate Change, Sustainable Development, and Energy: Future Perspectives for South Africa, Energy and Development Research Centre, University of Cape Town.

29. Dawson, B. and Spannagle, M., 2008. United Nations framework convention on climate change (Unfccc). In The complete guide to climate change (pp. 392-403). Routledge.

30. DeCola, P., Secretariat, W.M.O., 2017. An integrated global (GHG'S) information system (IGIS). World Meteorol. Org. Bull. 66 (1), 38–45.

31. Deleuil, T. (2012). The common but differentiated responsibilities principle recognises that while a united effort is required against climate change, all countries do not have the same capacity to contribute. Review of European Community & International Environmental Law, 21(2), 271-282. doi:10.1111/reel.12005

32. Department of Minerals and Energy (DME), (2002), South African National Energy Database, DME Statistics, Pretoria.

33. Donahue, J., Hendricks, L.A., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K. and Darrell, T., 2015. Long-term recurrent convolutional networks for visual recognition and description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp.2625-2634.

34. Liu, D., Guo, X. and Xiao, B., 2019. What causes the growth of global greenhouse gas emissions? Evidence from 40 countries. Science of the Total Environment, 661, pp.750-766. Available at: https://doi.org/10.1016/j.scitotenv.2019.01.197 [Accessed 5 September 2024].

35. Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Kadner, S., Minx, J., Brunner, S., Agrawala, S. and Baiocchi, G., 2014. Technical summary. In: Climate Change 2014: Mitigation of Climate Change. IPCC Working Group III Contribution to AR5. Cambridge: Cambridge University Press, pp.31-116.

36. Edenhofer, O., Kadner, S., von Stechow, C., Schwerhoff, G. and Luderer, G., 2014. Linking climate change mitigation research to sustainable development. In Handbook of Sustainable Development (pp. 476-499). Edward Elgar Publishing.

37. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical Evaluation of Gated

Recurrent Neural Networks on Sequence Modeling. CoRR abs/1412.3555. Published 2014. Available from: https://arxiv.org/abs/1412.3555.

38. Edenhofer, O., Jakob, M., Creutzig, F., Flachsland, C., Fuss, S., Kowarsch, M., Lessmann, K., Mattauch, L., Siegmeier, J. and Steckel, J.C., 2015. Closing the emission price gap. Global environmental change, 31, pp.132-143.

39. Engelbrecht F, Adegoke J, Bopape MJ, Naidoo M, Garland R, Thatcher M, et al. Projections of rapidly rising surface temperatures over Africa under low mitigation. Environ Res Lett. 2015;10(8):085004. doi:10.1088/1748-9326/10/8/085004.

40. Farge E. UN declares access to a clean environment a human right. Accessed on: November 05, 2021. Available from URL: https://www. reuters.com/business/environment/un-passes-resolution-makingclean-environment- access-humanright-2021-10-08/.

41. Graves, A., 2013. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.

42. G. Zhu, F. Porikli, and H. Li, "Robust visual tracking with deep convolutional neural network-based object proposals on pets," in IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, Las Vegas, NV, 2016, pp. 1265-1272.

43. GN 580 in GG 41689 of 8 June (hereinafter the Bill).

44. Gulia, S.; Nagendra, S.S.; Khare, M.; Khanna, I. Urban air quality management-A review. Atmos. Pollut. Res. 2015, 6, 286–304.

45. Liebenberg-Enslin, H. and Hurt, Q., 2009. Gauteng Province Air Quality Management Plan – Final Report. Technical Report No. 1. Gauteng

Department of Agriculture, Conservation and Environment

46. Hanto, J., Schroth, A., Krawielicki, L., Oei, P.Y. and Burton, J., 2022. The political economy of energy and climate policy in South Africa. In: P.Y. Oei, L. Krawielicki, A. Schroth and J. Burton, eds. The Political Economy of Coal. 1st ed. p.300.

47. Humby, T. (2018). The Thabametsi Case: Case No 65662/16 Earthlife Africa Johannesburg v Minister of Environmental Affairs. Journal of Environmental Law, 145, 145-157.

48. IPCC, 2021. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., et al., (Eds.). Cambridge: Cambridge University Press.

49. Janssens-Maenhout, G.; Crippa, M.; Guizzardi, D.; Dentener, F.; Muntean, M.; Pouliot, G.; Keating, T.; Zhang, Q.; Kurokawa, J.; Wankmüller, R.; et al. HTAP_v2.2: A mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution. Atmos. Chem. Phys. Discuss. 2015, 15, 11411–11432.

50. Fu, X., Huang, J., Ding, X., Liao, Y. and Paisley, J., 2017. Clearing the skies: A deep network architecture for single-image rain removal. IEEE Transactions on Image Processing, 26(6), pp.2944-2956.

51. Cho, K., van Merrienboer, B., Gulcehre, C., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1724–1734). Doha, Qatar.

52. Karpathy A, Fei-Fei L. Deep Visual-Semantic Alignments for Generating Image Descriptions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015:3128-3137. doi:10.1109/CVPR.2015.7298932.

53. Kaufman YJ, Tznr DC, Gordon HR, Nakajima T, Lenoble J, Frouins R, Grassl H, Herman BM, King MD, Teillet PM (1997) Passive remote sensing of tropospheric aerosol and atmospheric correction for theaerosol effect. J Geophys Res 102(D14):16815–16830.

54. Koneswaran, G., & Nierenberg, D. (2008). Global farm animal production and global warming: impacting and mitigating climate change. Environmental Health Perspectives, 116(5), 578-582.doi:10.1289/ehp.11034

55. Kruger AC, Nxumalo MP. Historical rainfall trends in South Africa: 1921–2015. Water SA. 2017 Apr 21; 43(2): 285–97.

56. Kumar, U. and Jain, V.K., 2010. ARIMA forecasting of ambient air pollutants (O 3, NO, NO 2 and CO). Stochastic Environmental Research and Risk Assessment, 24, pp.751-760.

57. Pan, L., Li, H., Li, W., Chen, X., Wu, G., & Du, Q. (2017). Discriminant analysis of hyperspectral imagery using fast kernel sparse and low-rank graph. IEEE Transactions on Geoscience and Remote Sensing, 55(11), 6085-6098.

58. Pak, D., Kim, H. and Lee, K., 2018. A CNN-LSTM model for predicting air pollution in Beijing. *Atmospheric Environment*, 193, pp.126-137. doi:10.1016/j.atmosenv.2018.09.053.

59. Leila, R. (2013). Climate change impacts on North African countries and

some Tunisian economic sectors. Journal of Agriculture and Environment for International Development, 107(1), 5-26.

60. Lelieveld, J., Klingmüller, K., Pozzer, A., Burnett, R. T., Haines, A., and Ramanathan, V. (2019). Effects of fossil fuel and total anthropogenic emission removal on public health and climate. Proceedings of the National Academy of Sciences, 116(15):7192–7197.

61. Lespinas, F., Wang, Y., Broquet, G., Bréon, F.M., Buchwitz, M., Reuter, M., Meijer, Y., Loescher, A., Janssens-Maenhout, G., Zheng, B., & Ciais, P. (2020). The potential of a constellation of low earth orbit satellite imagers to monitor worldwide fossil fuel CO2 emissions from large cities and point sources. Carbon Balance and Management, 15(1), 16.

62. Lin K-P, Pai P-F, Yang S-L. Forecasting concentrations of air pollutantsby logarithm support vector regression with immune algorithms. Applied Math Computer. 2011;217(12):5318–27.

63. Zhang, L., Zhang, L., Mou, X., & Zhang, D. (2011). FSIM: A feature similarity index for image quality assessment. IEEE Transactions on Image Processing, 20(8), 2378–2386.

64. Liousse C, Assamoi E, Criqui P, Granier C, Rosset R et al (2014) Explosive growth inAfrican combustion emissions from 2005 to 2030. Environmental Research Letters 9.https://doi.org/10.1088/1748-9326/9/3/035003.

65. Lipton, Z.C., Berkowitz, J. and Elkan, C., 2015. A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.

66. Ma, J.; Beirle, S.; Jin, J.; Shaiganfar, R.; Yan, P.; Wagner, T. Tropospheric

NO2 vertical column densities over Beijing: Results of the first three years of ground-based MAX-DOAS measurements (2008–2011) and satellite validation. Atmos. Chem. Phys. 2013, 13, 1547–1567.

67. Ranzato, M., Szlam, A., Bruna, J., Mathieu, M., Collobert, R., & Chopra, S. (2014). Video (language) modeling: A baseline for generative models of natural videos. arXiv preprint arXiv:1412.6604.

68. Ranzato, M., Szlam, A., Bruna, J., Mathieu, M., Collobert, R. and Chopra, S., 2014. Video (language) modeling: a baseline for generative models of natural videos. arXiv preprint arXiv:1412.6604.

69. M.D., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C., et al. (2014). Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press.

70. MacFaul, L. (2007). Monitoring Greenhouse Gases (GHGs). Bulletin of the American Meteorological Society, 88(8), 1171.

71. Matsueda, H., Moore, F., Morino, I., Park, S., Robinson, J., Roehl, C.M., Sawa, Y., Sherlock, V., Sweeney, C., Tanaka, T., & Zondlo, M.A. (2010). Calibration of the Total Carbon Column Observing Network using aircraft profile data. Atmospheric Measurement Techniques, 3, 1351–1362. https://doi.org/10.5194/amt-3-1351-2010.

72. Nyasulu, M., Haque, Md.M., Kumar, K.R., Banda, N., Ayugi, B., & Uddin, Md.J. (2021). Temporal patterns of remote-sensed tropospheric CO_2 and CH_4 over an urban site in Malawi, Southeast Africa: Implications for climate effects. Atmospheric Pollution Research, 12(3), 125-135.

73. Matthias Drusch, Umberto Del Bello, Sebastien Carlier, Olivier

Colin,Veronica Fernandez, Ferran Gascon, Bianca Hoersch, Claudia Isola, Paolo Laberinti, Philippe Martimort, et al. Sentinel-2: Esa's optical high-resolution mission for Games operational services. Remote sensing of Environment, 120:25–36, 2012.

74. Mbokodo I, Bopape MJ, Chikoore H, Engelbrecht F, Nethengwe N. Heatwaves in the future warmer climate of South Africa. Atmosphere. 2020 Jul; 11(7): 712.

75. Meijer, Y., Boesch, H., Bombelli, A., Brunner, D., Buchwitz, M., Ciais, P., Crisp, D., Engelen, R., Holmlund, K., Houweling, S., Janssens-Maenhout, G., Marshall, J., Nakajima, M., Pinty, B., Scholze, M., Bezy, J.-L., Drinkwater, M.R., Fehr, T., Fernandez, V., Loescher, A., & Nett, H. (2020). Copernicus CO_2 Monitoring Mission Requirements Document (MRD). European Space Agency, Earth and Mission Science Division, revision 3 (October 1, 2020). https://esamultimedia.esa.int/docs/EarthObservation/CO2M_MRD_v3.0_20201001_Issued.pdf.

76. Menut, L.; Bessagnet, B.; Khvorostyanov, D.; Beekmann, M.; Blond, N.; Colette, A.; Coll, I.; Curci, G.; Foret, G.; Hodzic, A.; et al. CHIMERE 2013: A model for regional atmospheric composition modelling. Geosci. Model Dev. 2013, 6, 981–1028.

77. Mishra, D. and Goyal, P., 2016. Neuro-Fuzzy approach to forecasting Ozone Episodes over the urban area of Delhi, India. Environmental Technology & Innovation, 5, pp.83-94.

78. Morgan, W.T., Allan, J.D., Bower, K., Highwood, E.J., Liu, D., McMeeking, G.R., Northway, M.J., Williams, P.I., Krejci, R. and Coe, H., 2010. Airborne measurements of the spatial distribution of aerosol chemical

composition across Europe and evolution of the organic fraction. Atmospheric Chemistry and Physics Discussions, 10, pp.4065-4083.

79. Ngaira, J. (2007). Impact of climate change on agriculture in Africa by 2030. Scientific Research and Essays, 2, 238–243.

80. Nkambule, N.P., & Blignaut, J.N. (2017). Externality costs of the coal-fuel cycle: The case of Kusile Power Station. South African Journal of Science, 113(9–10), 1–9.

81. O'Brien, D. M. and P. J. Rayner (2002) Global observations of the carbon budget 2. CO2 column from differential absorption of reflected sunlight in the 1.61 µm band of CO2, Journal of Geophysical Research, 107 (D18), ACH 6-1 – ACH 6-1.

82. Pacala, S., Breidenich, C., Brewer, P.G., Fung, I., Gunson, M.R., Heddle, G., Law, B., Marland, G., Paustian, K., Prather, M., Randerson, J.T., Tans, P., & Wofsy, S.C. (2010). Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements. In Proceedings of Advances in Neural Information Processing Systems (NIPS) (pp. 2565–2584).
83. National Research Council, Division on Earth, Life Studies, Board on Atmospheric Sciences and Committee on Methods for Estimating Greenhouse Gas Emissions. (2010). Verifying greenhouse gas emissions: methods to support international climate agreements. National Academies Press.

84. Pascanu, R., 2013. On the difficulty of training recurrent neural networks. arXiv preprint arXiv:1211.5063.

85. Radhakrishnan, M., Pathirana, A., Ashley, R., & Zevenbergen, C. (2017). Structuring Climate Adaptation through Multiple Perspectives: Framework and Case Study on Flood Risk Management. Water, 9(2), 129

86. Rao, A.N. and Chandrasena, N.R., 2022. The Need for Climate-Resilient Integrated Weed Management (CRIWM) under future Climate Change. Weeds-Journal of the Asian-Pacific Weed Science Society, 4(2), pp.1-20.Eial ir.

87. Revi, A., Satterthwaite, D., Aragón-Durand, F., Corfee-Morlot, J., Kiunsi, R.B.R., Pelling, M., Roberts, D.C., & Solecki, W. (2014). Urban Areas. In Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change; Field, C.B., Barros, V.R., Dokken, D.J., Mach, K.J., Mastrandrea, M.D., et al. (Eds.). Cambridge: Cambridge University Press.

88. Robinson, D.; Lloyd, C.; McKinley, J. Increasing the accuracy of nitrogen dioxide (NO2) pollution mapping using geographically weighted regression (GWR) and geostatistics. Int. J. Appl. Earth Obs. Geoinf. 2013, 21, 374–383.

89. Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137-1149.

90. Salman, A.G., Heryadi, Y., Abdurahman, E., & Suparta, W. (2018). Single layer & multi-layer long short-term memory (LSTM) model with intermediate variables for weather forecasting. Procedia Computer Science, 135, 89–98

91. Section 1 of the National Environmental Management Act: Air QualityAct 39 of 2004.

92. Seo, J.; Kim, J.Y.; Youn, D.; Lee, J.Y.; Kim, H.; Lim, Y.B.; Kim, Y.; Jin, H.C. On theMultiday Haze in the Asian Continental Outflow: The Important Role of Synoptic Conditions Combined with Regional and Local Sources. Atmos. Chem. Phys. 2017, 17, 9311–9332.

93. Srivastava, N., Mansimov, E. and Salakhudinov, R., 2015, June. Unsupervised learning of video representations using lstms. In International conference on machine learning (pp. 843-852). PMLR.

94. Shpak, N., Ohinok, S., Kulyniak, I., Sroka, W., Fedun, Y., Ginevičius, R. and Cygler, J., 2022. CO2 emissions and macroeconomic indicators: Analysis of the most polluted regions in the world. Energies, 15(8), p.2928.

95. Shivangi Saxena Somvanshi, Aditi Vashisht, Umesh Chandra, and Geetanjali Kaushik. Delhi Air Pollution Modeling Using Remote Sensing Technique. Amity Institute of Environmental Sciences, Amity University, Sector – 125, Noida, UP, India.

96. Shobande, O. A. (2020). The effects of energy use on infant mortality rates in Africa. Environmental and Sustainability Indicators, 5, 100015.

97. Sifakis, N., & Deschamps, P.Y. (1992). Mapping of air pollution using SPOT satellite data. Photogrammetric Engineering and Remote Sensing, 58(10), 1433–1437.

98. Tsai Y-T, Zeng Y-R, Chang Y-S. Air pollution forecasting using rnn with lstm. In: 2018 IEEE 16th Intl Conf on dependable, autonomic and secure computing, 16th intl conf on pervasive intelligence and computing, 4th intl conf on big data intelligence and computing and cyber science and technology congress(DASC/PiCom/DataCom/CyberSciTech). IEEE; 2018, p. 1074–9.

99. Ullah, A., Raza, K., Nadeem, M., Mehmood, U., Agyekum, E.B., Elnaggar, M.F., Agbozo, E. and Kamel, S., 2022. Does globalization cause greenhouse

gas emissions in Pakistan? A promise to enlighten the value of environmental quality. International Journal of Environmental Research and Public Health, 19(14), p.8678.

100. UN, E., 2019. Government Survey 2012: E-Government for the People. United Nations (UN) (2012).

101. Vardoulakis, S.; Solazzo, E.; Lumbreras, J. Intra-urban and street scale variability of BTEX, NO2 and O3 in Birmingham, UK: Implications for exposure assessment. Atmos. Environ. 2011, 45, 5069–5078.

102. variables for weather forecasting. Procedia Comput Sci. 2018;135:89–98.

103. Varon, D. J., Jervis, D., Pandey, S., Gallardo, S. L., Balasus, N., Yang, L. H., & Jacob, D. J. (2023). Quantifying NOx point sources with Landsat and Sentinel-2 satellite observations of NO2 plumes. Journal of Geophysical Research: Atmospheres, 128(4), e2022JD037668. doi:10.1029/2022JD037668.

104. Wang P, Liu Y, Qin Z, Zhang G. A novel hybrid forecasting model forpm10 and so2daily concentrations. Sci Tot Environ. 2015;505:1202– 12.

105. White Paper 8.

106. White Paper 8.

107. White Paper 9. The effects of climate change on South Africa particularly are of concern. The effects would be felt across important sectors such as water and agriculture. See Van der Bank and Karsten 2020 Air, Soil and Water Research 1. Alsosee the discussion in part 4 below and

Niang et al "Africa" 1209.

108. Wiser, R., and Zhang, X. (2014). Energy Systems. In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

109. Wunch, D., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Stephens, B. B., Fischer, M. L., Uchino, O., Abshire, J. B., Bernath, P., Biraud, S. C., Blavier, J.-F. L., Boone, C., Bowman, K. P., Browell, E. V., Campos, T., Connor, B. J., Daube, B. C., Deutscher, N. M., … & others. (2011). A method for evaluating bias in global measurements of CO2 total columns from space. Atmospheric Chemistry and Physics, 11(23), 12317-12337.

110. Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K. and Woo, W.C., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in Neural Information Processing Systems, 28, pp.802-810.

111. Xing, J.; Mathur, R.; Pleim, J.; Hogrefe, C.; Gan, C.-M.; Wong, D.C.; Wei, C.; Gilliam, R.; Pouliot, G. Observations and modeling of air quality trends over 1990–2010 across the Northern Hemisphere: China, the United States and Europe. Atmos. Chem. Phys. Discuss. 2015, 15, 2723–2747.

112. Xu, Z., Sheffield, P., Hu, W., Su, H., Yu, W., Qi, X., and Tong, S. (2012). Climate change and children’s health—a call for research on what works to protect children. International Journal of Environmental Research and Public Health, 9(9):3298–3316.

113. Chen, Y., Lin, Z., Zhao, X., Wang, G., & Gu, Y. (2014). Deep learning-based classification of hyperspectral data. IEEE Journal of Selected

Topics in Applied Earth Observations and Remote Sensing, 7(6), 2094-2107.

114. Y., Goldfarb, L., Gomis, M.I., et al., Eds.; Cambridge University Press:Cambridge, UK, 2021; in press.

115. Yang, Z. D., Bi, Y. M., Wang, Q., Liu, C. B., Gu, S. Y., Zheng, Y. Q.,Lin, C., Yin, Z.S., and Tian, L. F.: Inflight Performance of the TanSat Atmospheric Co2Grating Spectrometer, IEEE T. Geosci. Remote, 58, 4691–4703, 2020.

116. Sun, Y., Yin, H., Wang, W., Shan, C., Notholt, J., Palm, M., Liu, K., Chen, Z., & Liu, C. (2022). Monitoring GHGs in China: Status and perspective. Atmospheric Measurement Techniques, 15, 4819–4834.

117. Yu R, Yang Y, Yang L, Han G, Move OA. Raq-a random forest approach for predicting air quality in urban sensing systems. Sensors. 2016;16(1):86.

118. Wang, Z., Bovik, A.C., Sheikh, H.R., & Simoncelli, E.P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600–612.

119. Woo, W.C. and Wong, W.K., 2014. Application of optical flow techniques to rainfall nowcasting. In: 27th Conference on Severe Local Storms, 3-7 November 2014, Madison, Wisconsin.

120. Zhou, S., Shen, W., Zeng, D., Fang, M., Wei, Y., & Zhang, Z. (2016). Spatio-temporal convolutional neural networks for anomaly detection and localization in crowded scenes. Signal Processing: Image Communication, 47, 358–368.

121. Zhou, S., Zhang, X., Chu, S., Zhang, T., & Wang, J. (2020). Research on remote sensing image carbon emission monitoring based on

deep learning. Tallit Science and Technology Co., Ltd., Hefei 230088, China; Anhui Jiyuan Testing Technology Co., Ltd., Hefei 230088, China; Forestry College, Beijing Forestry University, Beijing 100091, China.

122. Ziervogel, G., New, M., Archer van Garderen, E., Midgley, G., and Taylor, A., 2020. Climate change impacts in Southern Africa: Implications for adaptation. Annual Review of Environment and Resources, 45, pp.367-396

123. Ziervogel, G., New, M., Archer van Garderen, E., Midgley, G., Taylor, A., Hamann, R., Stuart-Hill, S., Myers, J. and Warburton, M., 2014. Climate change impacts and adaptation in South Africa. Wiley Interdisciplinary Reviews: Climate Change, 5(5), pp.605-620.