Universidade do Minho
Escola de Engenharia

João Delfim Da Cruz Pereira

Using Deep Learning Models for Image Classification: A case study in water pollution

October, 2025

Universidade do Minho
Escola de Engenharia

João Delfim Da Cruz Pereira

Using Deep Learning Models for Image Classification: A case study in water pollution

Master's Dissertation
Master's in Informatics Engineering

Work supervised by
Paulo Jorge Freitas De Oliveira Novais

October, 2025

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the Universidade do Minho.

Braga , 14th October 2025

(Location) (Date)

Assinado por: **João Delfim da Cruz Pereira**
Num. de Identificação: 14969149
Data: 2025.10.14 15:51:34+01'00'

(João Delfim Da Cruz Pereira)

Acknowledgements

First and foremost, I want to express my gratitude to my supervisor, Professor Paulo Novais, for the mentorship of my dissertation, which provided the conditions and expertise to fund this work. Without his orientation, this wouldn't have been possible.

I also want to thank Professor Pedro Oliveira for his continuous support and help throughout the dissertation and the brainstorming sessions for the practical implementation, to bring this project to fruition.

In addition, I want to express my gratitude to Professor Gilberto Martins for providing the in-situ dataset for the two study locations. Therefore, without this data, this study would not have been possible.

An additional thanks goes to the Centro Algoritmi for accepting and welcoming me into the AIM4Water project. This also extends to the University of Minho for allowing me to pursue the Master's in Informatics Engineering programme.

Another special and heartfelt acknowledgement is reserved for my family, especially my mother, father, and brothers, for supporting my personal, academic, and professional paths throughout my life. Without their unconditional support and love, this wouldn't have been possible, even though they're currently far away. This work is as much yours as it is mine.

To my dear friends from Luxembourg and Portugal, who supported me on various occasions during decompression and helped me regain the energy needed for this dissertation.

This work is partially financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project 2022.06822.PTDC (https://doi.org/10.54499/2022.06822.PTDC.

Abstract

The increasing availability of satellite imagery from missions like Sentinel-2, combined with advances in Deep Learning (DL), presents a compelling opportunity to develop scalable, automated, and low-cost systems for monitoring water quality over large areas.

Water pollution poses a significant threat to ecosystems and public health. Yet, traditional in-situ monitoring methods are often costly, infrequent, and limited in spatial coverage. Therefore, this dissertation proposes and validates a DL framework for classifying Lake Water Quality using a multimodal approach that integrates in-situ data with Sentinel-2 satellite imagery for two Azorean Lakes. The framework systematically evaluates state-of-the-art architectures, including Convolutional Neural Networks (CNNs) and Transformers. A robust preprocessing pipeline was developed to handle cloud cover and calculate spectral indices, while a Water Quality Index (WQI) was engineered from ground-truth measurements to serve as labels. The study rigorously compares model performance under 3 distinct paradigms: training from scratch, training with extensive data augmentation, and applying Transfer Learning (TL) with pre-trained models, all optimised through a systematic hyperparameter search.

The results demonstrate that data augmentation is a critical strategy, consistently yielding the highest classification performance across all models, with F1 scores frequently exceeding 0.979. While TL offers a compelling trade-off between performance and computational cost, its success is highly context-dependent, presenting a "high-risk, high-reward" scenario. Ultimately, this dissertation confirms the feasibility of using DL and Remote Sensing (RS) to build a robust, scalable, and automated framework for environmental monitoring. This solution provides a valuable tool for sustainable water resource management, transforming raw satellite data into actionable insights to preserve aquatic ecosystems.

Keywords: Deep Learning, Image Classification, Remote Sensing, Water Quality Index.

Resumo

A crescente disponibilidade de imagens de satélite de missões como a Sentinel-2, em combinação com os avanços em DL, apresenta uma oportunidade promissora para o desenvolvimento de sistemas escaláveis, automatizados e de baixo custo para a monitorização da qualidade da água em grandes áreas.

A poluição da água representa uma ameaça significativa para os ecossistemas e para a saúde pública, mas os métodos tradicionais de monitorização in situ são frequentemente dispendiosos, pouco frequentes e com uma cobertura espacial limitada. Por isso, esta dissertação propõe e valida uma estrutura DL para classificar a qualidade da água dos lagos utilizando uma abordagem multimodal que integra dados in situ com imagens de satélite Sentinel-2 para dois lagos açorianos. A estrutura avalia sistematicamente arquiteturas presentes na literatura, incluindo CNNs e Transformers. Nesse sentido, foi desenvolvido um pipeline de pré-processamento robusto para lidar com a cobertura de nuvens e calcular índices espectrais. Além disso, um WQI foi concebido a partir de medições de *ground-truth* para servir como *labels*. O estudo compara rigorosamente o desempenho do modelo sob três paradigmas distintos: treino de raiz, treino com *data augmentation* e aplicação de TL com modelos pré-treinados, todos otimizados através de uma pesquisa sistemática de hiperparâmetros.

Os resultados demonstram que a técnica de *data augmentation* é uma estratégia fundamental, pois produz consistentemente o melhor desempenho de classificação em todos os modelos, com pontuações F1 frequentemente superiores a 0,979. Embora os modelos pré-treinados ofereçam um equilíbrio convincente entre desempenho e custo computacional, o seu sucesso é altamente dependente do contexto, apresentando um cenário de "alto risco e alta recompensa". Em última análise, esta dissertação confirma a viabilidade da utilização de DL e RS para construir uma estrutura robusta, escalável e automatizada para a monitorização ambiental. Esta solução fornece uma ferramenta valiosa para a gestão sustentável dos recursos hídricos, transformando dados brutos de satélite em insights acionáveis para preservar os ecossistemas aquáticos.

Palavras-chave: Aprendizagem Profunda, Classificação de Imagens, Sensorização Remota, Índice de Qualidade da Água.

Contents

List of Figures

List of Tables

Acronyms

AI	Artificial Intelligence
AlexNet	Alexander Network
ANN	Artificial Neural Network
API	Application Programming Interface
AT	Atmospheric Temperature
BERT	Bidirectional Encoder Representations from Transformers
BGA-PC	Blue-Green Algae Phycocyanin
BMP	Best Management Practice
BOD	Biochemical Oxygen Demand
CDOM	Coloured Dissolved Organic Matter
Chl-a	Chlorophyll-a
CLD	Cloud Band
CNN	Convolutional Neural Network
CO_2	Carbon Dioxide
COD	Chemical Oxygen Demand
CPU	Central Processing Unit
CSV	Comma Separated Values
CWA	Clean Water Act
DenseNet	Dense Convolutional Network
DL	Deep Learning
DNN	Deep Neural Network
DO	Dissolved Oxygen
DT	Decision Tree
E. coli	Escherichia Coli
EPR	Extended Producer Responsibility

EQR	Ecological Quality Ratios
ERTS	Earth Resources Technology Satellite
ESA	European Space Agency
ETM	Enhanced Thematic Mapper
ETM+	Enhanced Thematic Mapper Plus
EU	European Union
F1-score	Harmonic Mean of the Precision and Recall
FAI	Floating Algae Index
FCL	Fully Connected Layers
fDOM	Fluorescent Dissolved Organic Matter
FNN	Feedforward Neural Network
GAN	Generative Adversarial Network
GELU	Gaussian Error Linear Unit
GMES	Global Monitoring for Environment and Security
GNN	Graph Neural Networks
GOS	Global Observing System
GPT	Generative Pre-trained
GPU	Graphical Processing Unit
HABs	Harmful Algal Blooms
HSV	Hue-Saturation Value
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IR	Infrared
IS	Intelligent System
LIME	Local Interpretable Model-Agnostic Explanation
LSTM	Long Short-Term Memory
LULC	Land Use and Land Cover
MAD	Median Absolute Deviation
MAPE	Mean Absolute Percentage Error
MAS	Multi-Agent System
MetOp-SG	Meteorological Operational Satellite - Second Generation

MHA	Multi-Head Attention Mechansim
ML	Machine Learning
MLP	Multilayer Perceptron
MLR	Multiple Linear Regression
MRI	Magnetic Resonance Imaging
MSE	Mean Square Error
MSS	Multispectral Scanner
MTG-S	Meteosat Third Generation Sounders
NASA	National Aeronautics and Space Administration
NBRI	Normalized Burned Ratio Index
NDCI	Normalized Difference Chlorophyll Index
NDSI	Normalized Difference Snow Index
NDTI	Normalized Difference Turbidity Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NGO	Nongovernamental Organization
NH_3	Ammonia
NH_3-N	Ammonia Nitrogen
NIR	Near-Infrared
NLP	Natural Language Processing
NO_3	Nitrates
NO_x	Nitrogen Oxides
NPS	Non-Point Source
O_2	Oxygen
OLI	Operational Land Imager
pH	Potential of Hydrogen
PLS	Partial Least Squares
PP	Precipitation
PS	Point Source
PSO	Particle Swarm Optimization
R^2	Coefficient of Determination
ReLU	Rectified Linear Unit

ResNet	Residual network
ResNeXt	Residual Networks with Aggregated Transformations
RGB	Red-Green-Blue
RH	Relative Humidity
RL	Reinforcement Learning
RMSE	Root-Mean Square Error
RMSProp	Root Mean Square Propagation
ROI	Region of Interest
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SDG	Sustainable Development Goal
SHAP	Shapley Additive Explanations
SL	Supervised Learning
SMOTE	Synthetic Minority Over-sampling Technique
SO_2	Sulphur Dioxide
STFT	Short-Time Fourier Transform
STP	Sewage Treatment Plant
SVM	Support Vector Machine
SVR	Support Vector Regression
Swin	Shifted Windows Transformer
Swin-MFINet	Swin transformer based multi-feature integration network
SWIR	Short-Wave Infrared
Tanh	Tangent
TGS	Tomographic Gamma Scanning
TIF	Tagged Image File
TIRS	Thermal Infrared Sensor
TL	Transfer Learning
TM	Thematic Mapper
TN	Total Nitrogen
TP	Total Phosphorus
TSS	Total Suspended Solids
UL	Unsupervised Learning
UN	United Nations

UNFCCC	United Nations Framework Convention on Climate Change
USGS	United States Geological Survey
VGGNet	Visual Geometry Group Network
ViT	Visual Transformer
WGS	World Geodetic System
WMO	World Meteorological Organization
WQI	Water Quality Index
WS	Wind Speed
WT	Water Temperature
WWTP	Wastewater Treatment Plant
XAI	Explainable Artificial Intelligence

Chapter 1

Introduction

Water pollution is one of the most pressing environmental challenges of our time, threatening ecosystems, human health, and sustainable development. The increasing prevalence of pollutants, including chemicals, plastics, and organic waste, in water bodies calls for innovative approaches to monitoring and mitigation. Recent advances in Artificial Intelligence (AI), particularly in image-based analysis, have opened new pathways for studying and managing water pollution. AI-powered techniques, using satellite imagery and advanced computer vision models, enable the detection, quantification, and analysis of pollutants with unprecedented precision and efficiency. This dissertation delves into the intersection of AI and environmental science, focusing on how modern image classification and semantic segmentation techniques can be applied to monitor water quality, detect harmful algal blooms, and identify pollution hotspots. By integrating cutting-edge AI methodologies, this work aims to develop scalable, automated tools to better understand and address water pollution.

1.1 Background & Motivation

This section aims to discuss the background and motivation of this dissertation. Water pollution is a critical global issue, and traditional monitoring methods are often limited in scope and efficiency. Recent advancements in AI and image analysis offer innovative solutions that enable automated, large-scale monitoring of water quality. This dissertation explores how these technologies can address key challenges in understanding and managing water pollution.

1.1.1 Intelligent Systems

Recent advances in Intelligent System (IS), including Machine Learning (ML), DL, and Reinforcement Learning (RL), have revolutionised how we analyse, interpret, and interact with complex data. These technologies enable computers to learn patterns, make predictions, and adapt their behaviour from data inputs that can be numerical, textual, or image-based [38]. By emulating aspects of human intelligence, IS can process vast amounts of information far beyond human capability, uncovering relationships and insights that were previously hidden. The use of RS enables the acquisition of much more image-based data of our world through satellite imagery. These can be used to learn more about our world and, therefore, serve as a monitoring tool [35]. As a result, IS serves as a crucial tool for addressing multifaceted challenges that span disciplines, from economics and healthcare to environmental science, enabling us, in conjunction with RS, to make better decisions that affect our world and society [71].

In our current times, environmental sustainability is a fundamental challenge that society faces. It is a part of the three big E's, which are economy, equality and ecology, as highlighted in Figure 1.1 [109]. The first one concerns how to build a thriving economy in the long run. The second one focuses on equality within the population, to maintain equilibrium so that everyone can benefit, without anyone having all the resources while others go without basic needs. The last one addresses how to maintain an ecosystem that is respected and still allows us to profit from it while it replenishes itself. If we combine all of these factors in a balanced way, this is called sustainability [14].

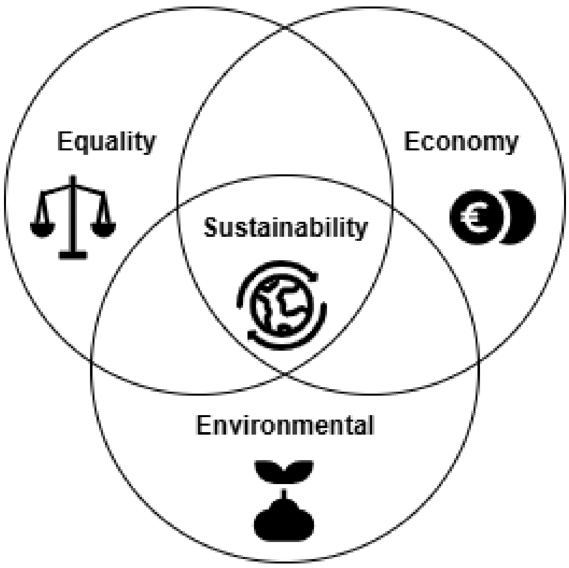

Figure 1.1: Three Big E's. Adapted from [8]

When discussing sustainability, the economic side plays a significant role in determining how raw materials and products are manufactured [14]. For example, if there is a cheaper way to manufacture the same product for less but without any way to recycle it later, companies will choose cost-effective ways of manufacturing over sustainability. Therefore, there are supranational entities that impose laws and policies that the public and private sectors need to respect. The European Union (EU) have been at the forefront of sustainability and have enacted many laws and policies that member states must adopt [109].

When discussing sustainability, equality serves as a cornerstone, ensuring that development benefits are distributed fairly across all segments of society and promoting inclusivity and social justice. This involves addressing disparities in access to resources, economic opportunities, and decision-making power. For instance, policies that prioritise equality may focus on ensuring fair wages, equitable access to education, or addressing gender gaps in employment. Supranational entities and organisations such as the glspl un emphasise equality through frameworks like the Sustainable Development Goals (SDGs) [41], particularly Goal 10, which aims to reduce inequalities within and among countries. By embedding equality into sustainability initiatives, societies seek to create systems in which all individuals can fully participate in economic, social, and environmental progress [109].

The environmental pillar of sustainability emphasises the stewardship of natural resources to maintain ecological balance and preserve biodiversity for future generations [14]. This involves practices that prioritise renewable energy, resource conservation, and ecosystem protection. For example, sustainable practices might include adopting clean energy technologies or implementing laws to protect endangered species and fragile ecosystems. Supranational organisations, such as the United Nations (UN) and the EU, actively promote environmental sustainability through frameworks like the Paris Agreement and the EU Green Deal, which encourage member states to reduce carbon emissions and invest in sustainable infrastructure. This global focus underscores the importance of the environment as a foundation for sustaining all life forms and maintaining planetary health [109].

To assess the environmental aspect of sustainability, robust systems must be in place to monitor and evaluate the effectiveness of policies and initiatives aimed at environmental protection [9]. Such systems provide critical data that enable decision-makers to understand environmental trends and measure progress over time. For example, the Copernicus Programme, led by the EU, is one of the most advanced Earth observation systems globally [138]. It uses satellite imagery and ground-based sensors to monitor environmental parameters, including air and water quality and land-use changes. Copernicus supports policy implementation by offering insights into climate trends, deforestation, and pollution levels, thereby ensuring that sustainability goals are met.

Similarly, the Global Observing System (GOS), coordinated by the World Meteorological Organization (WMO) [17], provides comprehensive monitoring of weather, climate, and water. These systems, combined with frameworks such as the United Nations Framework Convention on Climate Change (UNFCCC) reporting mechanisms, allow governments and organisations to track whether environmental policies, such as those targeting carbon reduction or biodiversity preservation, are achieving their desired outcomes. By using such monitoring tools, the impact of sustainability efforts becomes transparent and quantifiable, guiding further actions to improve ecological health [75].

1.1.2 Problems

When discussing environmental sustainability, it is essential to highlight the problems it faces today. Therefore, the upcoming sections describe the issues.

1.1.2.1 Climate Change

A major environmental issue is climate change, characterised by rising global average temperatures, intense droughts, water scarcity, severe fires, rising sea levels, flooding, melting polar ice, catastrophic storms, and declining biodiversity [9]. In reality, climate change boils down to rising temperatures, as figured in Figure 1.2, but the Earth is an interconnected system; if there is a change in one area, it can affect other areas.

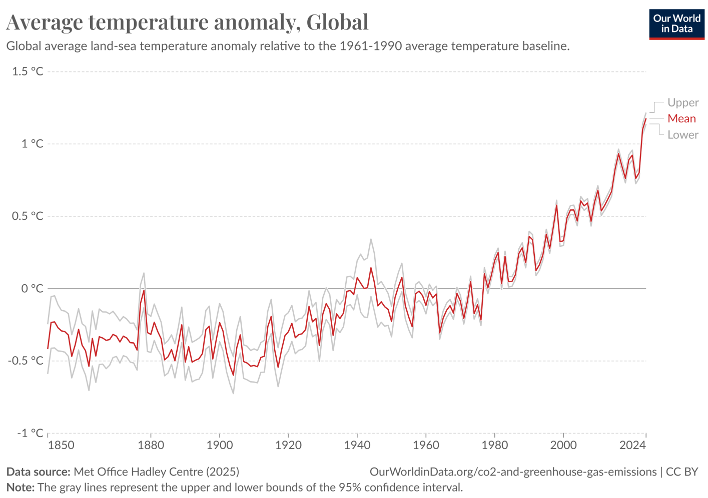

Figure 1.2: Temperature anomaly chart. Retrieved from [53]

When temperatures rise, the Earth gets significantly warmer, and all the ice, such as Antarctica and Greenland, starts to melt. All the melted ice turns into water and needs to go somewhere; in this case, it flows into the oceans, raising sea levels. These effects, then, have a direct impact on us humans. Countries and territories that are close to or below sea level could be flooded, such as the Netherlands or coastal regions. This will affect the lives of people living in those regions and destroy their property and belongings. The rising temperatures are connected to the amount of greenhouse gases emitted, such as Carbon Dioxide (CO_2), shown in Figure 1.3.

Other regions that already suffer from severe droughts and water scarcity will suffer even more. Rising sea levels, weather patterns, currents, and ocean water circulation also change. For Europe, a critical current is called the Gulf Stream. This stream is a warm-water current that originates in the Gulf of Mexico, flows up along Florida's coast, and veers eastward toward Europe, bringing warm water. When it reaches Europe, it can raise temperatures by 2 to 3 °C [105]. However, due to changes in ocean currents, this stream can be severely affected, leaving Europe without these phenomena.

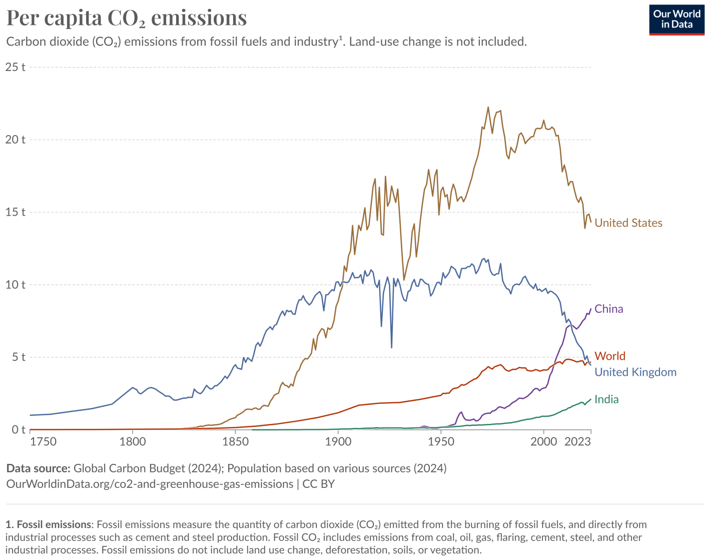

Figure 1.3: Emissions chart. Retrieved from [54]

1.1.2.2 Excessive Resource Consumption

One of the significant challenges regarding sustainability lies in how we consume natural resources and the rate at which these resources are replenished, commonly referred to as the replacement rate [123]. Figure 1.4 illustrates the consumption trends of various materials from 1970 to 2010, revealing a steady and alarming increase year after year. This pattern of consumption underscores a fundamental issue: resources, whether renewable or non-renewable, require time to regenerate naturally. Every resource has a specific replenishment cycle, but when we consume it faster than it can replenish itself, we disrupt that cycle. The imbalance between consumption and replenishment leads to resource depletion, as the demand outweighs the Earth's capacity to restore them. Over time, this excessive consumption depletes finite resources, creating a situation where, without intervention, certain materials may no longer be available to future generations. The concept of excessive consumption reflects this unsustainable approach, highlighting the urgency of addressing how we manage and utilise our resources to ensure long-term ecological balance [127].

1.1.2.3 Environmental Contamination

Another pressing issue tied to environmental sustainability is the introduction of harmful substances into ecosystems, which disrupts their balance and can have far-reaching consequences [136]. The overuse of chemicals such as pesticides, fertilisers, and industrial effluents in agricultural and urban settings often leads to these substances leaching into water sources and surrounding habitats. This not only harms aquatic and terrestrial species but also contaminates resources vital to human survival, such as drinking water and soil used for crops. Furthermore, microplastics and synthetic materials have found their way into every corner of the planet, from the deepest oceans to the remotest mountaintops, and have integrated

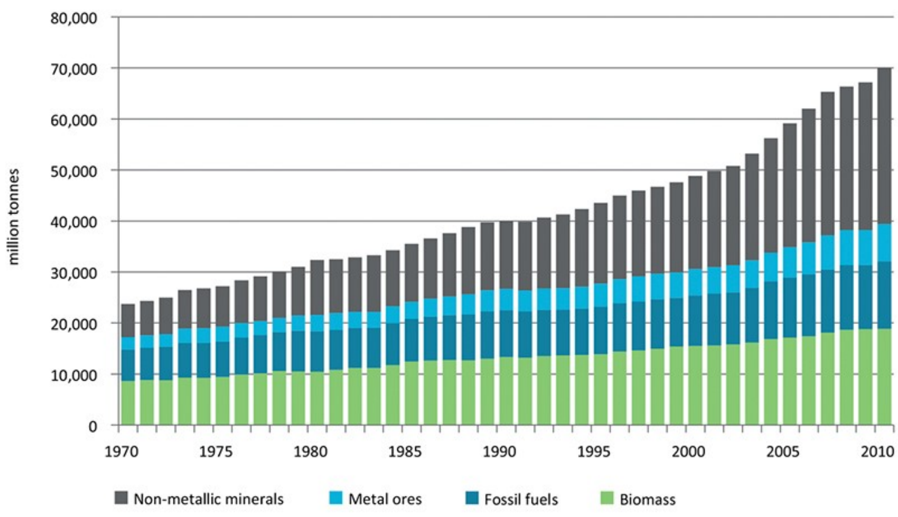

Figure 1.4: Resource Consumption chart. Adapted from [123]

into the food chain, impacting human health and biodiversity [69].

The connection to excessive resource consumption becomes evident here: many of these harmful materials stem from unsustainable manufacturing processes and over-reliance on non-biodegradable products. Similarly, climate change, driven by rising temperatures, amplifies the effects of environmental contamination, altering how ecosystems absorb and process these substances [136].

1.1.3 Proposed Solutions

The following sections outline the proposed solutions to address the challenges discussed above. These solutions focus on practical strategies and innovative approaches to promote sustainability, resource efficiency, and environmental responsibility.

One of the primary approaches to mitigating climate change is reducing carbon emissions through global agreements and initiatives, such as the Paris Agreement [40]. This treaty aims to limit global temperature rise to below 2 °C, with efforts to restrict it even lower to 1.5 °C. Achieving this requires nations to commit to reducing greenhouse gas emissions by adopting renewable energy sources, improving energy efficiency, and transitioning to electric mobility. The growth in renewable energy production is visualised through Figure 1.5, especially in wind and solar energy. Furthermore, reforestation initiatives and carbon capture technologies can significantly contribute to sequestering atmospheric CO_2. Governments, industries, and individuals must collaborate on sustainable practices to achieve these goals effectively.

To address the challenge of excessive resource consumption, adopting the principles of a circular economy is crucial [126]. A circular economy aims to create a closed-loop system in which products, materials, and resources are kept in use for as long as possible, minimising waste and reducing reliance

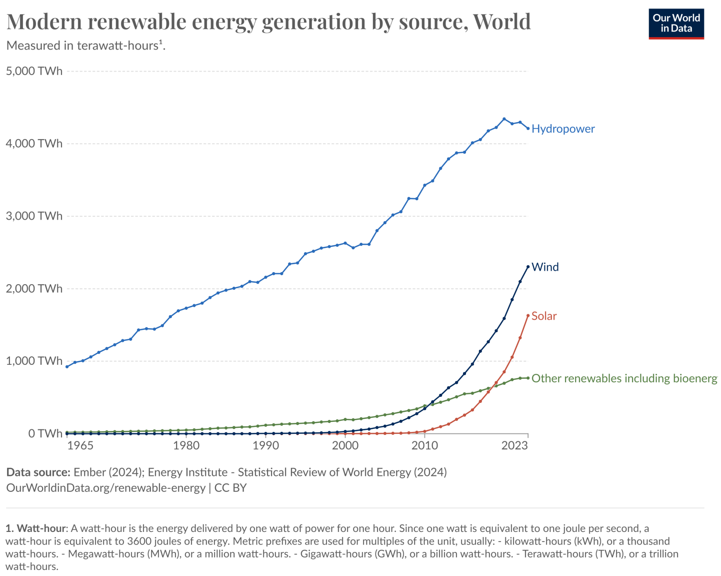

Figure 1.5: Renewable Energy Production chart. Retrieved from [115]

on finite resources. This concept stands in stark contrast to the traditional linear economy, which follows a “take-make-dispose” model. The circular economy encourages the design of products that can be reused, repaired, refurbished, or recycled, ensuring that the value of materials is preserved throughout their lifecycle [41].

One key aspect of the circular economy is the emphasis on sustainable product design [122]. Manufacturers are encouraged to develop products with a longer lifespan, using materials that are easier to recycle or repurpose. This not only reduces waste but also lessens the environmental impact of raw material extraction and the energy-intensive processes associated with production. For example, modular product designs, which allow components to be easily replaced or upgraded, can significantly extend a product’s life, reducing the need for disposal and new product manufacturing [41].

Another crucial component is the efficient use of resources. Circular economy practices promote reducing resource consumption through strategies such as resource recovery and material optimisation [118]. Industries can adopt processes that minimise waste generation and energy consumption, focusing on producing more with fewer resources. Additionally, by improving the efficiency of manufacturing and supply chains, companies can reduce unnecessary waste and their carbon footprint [41].

Governments play a vital role in supporting the transition to a circular economy [126]. Policy measures, such as incentivising industries to adopt sustainable production processes, can encourage businesses to invest in eco-friendly technologies and practices. Governments can also introduce tax benefits or subsidies for companies that prioritise sustainability, making it financially attractive for industries to adopt circular economy principles. Furthermore, Extended Producer Responsibility (EPR) regulations can require manufacturers to take responsibility for the disposal and recycling of their products, thus promoting environmentally responsible end-of-life management [41].

Public awareness is equally essential for the success of a circular economy. Educational campaigns that inform citizens about the benefits of responsible consumption can lead to shifts in consumer behaviour [118]. By fostering an understanding of the environmental and economic advantages of reusing, recycling, and repairing products, individuals can make more informed choices in their everyday lives. Community initiatives, such as repair workshops and recycling programs, can further engage consumers in the circular economy, empowering them to contribute to waste reduction and resource conservation [41].

The transition to a circular economy also opens new opportunities for innovation and job creation, as illustrated in Figure **??**. Industries focused on repair, refurbishment, recycling, and re-manufacturing are likely to see significant growth as demand for sustainable practices increases. This shift has the potential not only to reduce environmental harm but also to create a more resilient and sustainable economy that is less dependent on resource extraction and better able to adapt to changing environmental conditions.

Figure 1.6: Circular Economy illustration. Adapted from [126]

To combat environmental contamination, a multifaceted approach is necessary [9]. Implementing robust monitoring systems, such as the European Copernicus Earth Observation initiative, is crucial for identifying pollution sources and assessing environmental health in real-time. These systems enable governments and organisations to track air and water quality, greenhouse gas emissions, and ecosystem changes effectively [87].

In addition to monitoring, stricter regulations on waste management and chemical usage in industries and agriculture can significantly reduce pollution levels [48]. Encouraging the adoption of green technologies, such as advanced wastewater treatment plants, and promoting the use of biodegradable materials can further minimise harmful discharges into the environment [87].

Public participation and awareness are also vital. Governments can introduce incentive programs to promote recycling and reduce single-use plastics, while fostering community-led initiatives to clean local environments and preserve biodiversity [48]. Furthermore, advancing research into innovative solutions, such as microbial bioremediation and the development of materials that neutralise pollutants, offers long-term strategies for mitigating contamination [87].

By combining technological, regulatory, and community-based efforts, environmental contamination can be tackled comprehensively, paving the way for healthier ecosystems and sustainable development.

1.1.4 Pollution

Pollution refers to the introduction of harmful substances or contaminants into the natural environment, causing adverse effects on ecosystems, human health, and the planet's overall balance [56]. These pollutants can originate from various human activities, including industrial processes, agriculture, transportation, and urban development. Pollution affects air, water, and soil, each of which plays a critical role in sustaining life on Earth.

One significant way pollution affects us is through its impact on human health. For example, air pollution, caused by particulate matter and toxic gases, is linked to respiratory and cardiovascular diseases [12]. Water pollution, on the other hand, contaminates drinking water sources and harms aquatic ecosystems, leading to biodiversity loss and compromised food security. Soil pollution degrades agricultural land, reducing crop yields and increasing reliance on synthetic fertilisers [39].

Air Pollution comes from emissions from vehicles, factories, and power plants, which release harmful gases and particulates into the atmosphere. For instance, urban smog and acid rain result from these emissions [12].

Noise Pollution generated by excessive noise from transportation, construction, and industrial activities disrupts human well-being and animal habitats [12].

Light Pollution is the overuse of artificial lighting, which affects nocturnal wildlife and can disrupt human circadian rhythms [12].

Soil Pollution associated with the use of pesticides, improper waste disposal, and heavy metals leads to soil degradation [39].

Water Pollution tied to industrial waste, agricultural run-off, and untreated sewage pollute rivers, lakes, and oceans, harming aquatic ecosystems [56].

By addressing each of these pollution types, a healthier and more sustainable environment can be achieved; therefore, the focus of this study will be on the last one.

Water pollution is a pervasive and growing problem that affects both ecosystems and human health. It occurs when harmful substances, such as chemicals, pathogens, and debris, contaminate water sources, making them unsafe for consumption and disrupting aquatic ecosystems. These pollutants enter water bodies from various sources, and their effects are wide-ranging, impacting everything from biodiversity to human health and the economy [56].

When observing Figure 1.7, the consumption of water has been rising over the years, particularly rising from about 580 m^3 in 2020 to 1350 m^3 in 2023. Exponential growth does not come lightly, as it also increases the pressure to keep water sources clean for consumption.

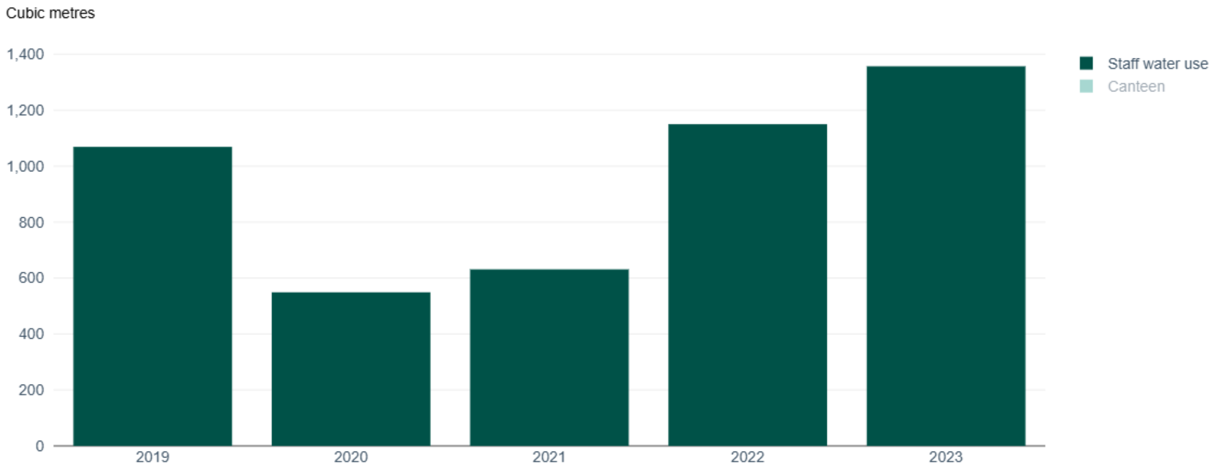

Figure 1.7: Water Consumption. Retrieved from [2]

The impacts of water pollution are both direct and indirect. From a human health perspective, polluted water is a leading cause of disease and death worldwide, particularly in low-income countries where sanitation infrastructure is inadequate [44]. Economically, water pollution impacts industries such as tourism, fishing, and agriculture. Coastal regions and rivers that suffer from pollution often see declines in tourism and in fish populations, severely affecting the livelihoods of people who depend on these resources. Ecosystems are similarly impacted, with species declining or going extinct due to the toxic effects of pollutants [19]. Wetlands, which act as natural water filters, are also being degraded, diminishing their ability to purify water and support biodiversity.

1.2 Objectives & Research Questions

The increasing availability of satellite imagery and advancements in DL have opened new possibilities for monitoring environmental issues such as water pollution. This dissertation aims to explore how RS data, combined with DL techniques, can contribute to the classification of polluted waters. By leveraging spectral indices and state-of-the-art image classification models, this study aims to develop a methodology that enhances the detection and analysis of water quality parameters. In addition to achieving the best possible outcomes from the DL model, a search strategy will be applied to extract more from said models.

This consolidation of a framework bridges the gap between environmental sustainability and scientific innovation, implements state-of-the-art AI and biological concepts, and applies them to real-world problems. Therefore, contributing to a more advanced society and the well-being of nature. It also aligns with the UN Sustainable Development Goals, which outline various sustainability goals. In our case, it contributes to Goal 6 for Clean Water and Sanitation, and Goal 13 for Climate Action.

To achieve this goal, the study is guided by the following objectives:

- RS: Retrieval of RS data, specifically Satellite imagery from Sentinel-2, including the creation of spectral indices;
- In-Situ Dataset: Obtainment and Preparation of data from laboratory analysis of the two study locations, with the merging of Meteorological Data;
- Image Classification: The merge of the two previous points into a unified dataset, to serve as input for the diverse DL models, for WQI classification.

Building on these objectives, this dissertation seeks to answer key research questions to assess the feasibility and impact of using RS and DL for water quality classification. The research questions are as follows:

RQ1: To what extent can freely available satellite RS provide accurate, low-cost, and scalable indicators of lake water quality compared with traditional in-situ monitoring?

RQ2: Which DL architectures and input representations deliver the best trade-off between classification performance and computational cost when classifying lake water quality?

RQ3: How do the spectral indices help to establish a link with the in-situ data that corresponds to different characteristics of the pollution in water?

By addressing these research questions, this study aims to bridge the gap between remote sensing and deep learning applications in water pollution monitoring, offering insights into how automated classification techniques can enhance environmental analysis.

1.3 Document Structure

The document is structured in the following way:

Chapter 1 begins by introducing the Background and Motivation, explaining what environmental sustainability is, the problems underlying it, and proposing solutions to the issues identified. Additionally, pollution will be presented, more specifically, water pollution. The objectives of this thesis will be given, along with three research questions.

Chapter 2 starts by laying out the research on state-of-the-art solutions already published, through articles, papers, and books. It describes the core ideas used throughout the experimentation phase of this dissertation, and also other technologies related to this field.

Chapter 3 presents the Materials and Methods, which includes how the study has been done. It includes how the data has been collected, the pre-processing steps, diagrams and images to visualise the data and the conception of the diverse models.

Chapter 4, specifically Experiments, includes the experimental setup, such as which hyperparameters have been used and how these have been searched. The inclusion of evaluation metrics, hardware information, and software versions.

Chapter 5 is called Results and Discussion, which states the results obtained of this thesis, through tables and explanations. At the end, an overview and comparison of all approaches will be included.

Chapter 6 is the last chapter, therefore being the conclusion, which concludes this thesis, through a summary of the work and future directions to keep improving this work.

Chapter 2

State of the Art

In this Chapter, we will examine how the research has been conducted to better understand the theoretical foundations, methodologies, and technologies that underlie this study. It starts by defining the research methodology in Section 2.1. Afterwards, in Section 2.2, we will explain what it means and how it impacts different water bodies, including their sources. Next up, in Section 2.3, we will define RS and the existing programmes, specifically satellites and their intricate details. In the subsequent Section 2.4, we highlight the basics of IS, including the different paradigms that lead to DL, and future directions. In Section 2.5, it will go much more in-depth into the architectures of different DL models. At the end, Section 2.6 includes 5 papers that have been chosen to perform a literature review, followed by Section 2.7, to conduct a critical analysis of the previous 5 papers, highlighting the positives, the negatives and the gaps.

2.1 Research Methodology

A necessary step is to identify and cite the relevant references to other studies and work that support the research. Therefore, a research methodology is needed to identify and select papers for this study from appropriate databases. First, the different purposes of this study needed to be defined: Environmental Sustainability, Pollution, Water Pollution, Intelligent Systems, Image Classification, DL Models, and RS. Second, the databases chosen were ResearchNet, Arxiv, ScienceDirect, IEEE Xplore, and Web of Science, all accessed via Google Scholar. At last, the need to use appropriate keywords were used, such as "Environmental Sustainability", "Environmental Sustainability Problems", "Environmental Sustainability Solutions", "Pollution", "Water Pollution", "Water Pollution Characteristics", "Water Pollution Problematic", "Water Pollution Examples", "Artificial Intelligence", "machine Learning", "Deep Learning", "Image Classification", "Deep Learning Models", "Satellite Imagery"and "Remote Sensing". The aforementioned keywords were used to find titles, abstracts, or keywords mentioned in the articles. The papers were written in English only, and the release years of the articles were limited to recent ones, beginning in 2019, except for the DL models. When finally choosing the articles, they were selected based on relevance and release year. The title, abstract, and keywords were analysed to assess their applicability to the research methodology; otherwise, when unsure, the introduction and conclusion were read. If all the criteria are met, the article has been fully read.

Purposes	**Keywords**	**Refinements**	**Selection Criteria**
- Environmental Sustainability	“Environmental Sustainability” “Environmental Sustainability Problems” “Environmental Sustainability Solutions” “Pollution”		
- Pollution	Water Pollution	Search on titles, abstracts, and keywords	Title, abstract, and keywords relevant
- Water Pollution	“Water Pollution Characteristics”	Only English Language	When there are doubts about its relevance, read the introduction and conclusion
- Intelligent Systems	“Water Pollution Problematic”	Limit to recent years (>2015)	If relevant, then the full article is read
- Image Classification	“Water Pollution Examples”	Look to previous years for specific articles	Stopping criterion: defined if the article does not match the defined purposes
- Deep Learning Models	“Artificial Intelligence” “Machine Learning” “Deep Learning” “Image Classification” “Deep Learning Models”		
Databases: - ResearchNet, Arxiv, ScienceDirect, IEEE Xplore, Web of Science			
Sort by: - Relevance, Release date			

Table 2.1: Research methodology strategy

2.2 Water Pollution

Water pollution is a pressing environmental issue that poses significant threats to ecosystems, human well-being, and planetary sustainability. It occurs when harmful substances enter water bodies, such as rivers, lakes, oceans, and groundwater, altering their natural composition and making them unsuitable for consumption, recreation, or supporting aquatic life [56]. Water pollution can occur from many sources.

Non-Point Source (NPS) Pollution is a widespread and diffuse form of water contamination that arises from multiple, unidentifiable sources. Unlike Point Source (PS) pollution, which can be traced to a specific location such as a factory discharge pipe or wastewater outlet, NPS pollution is carried by run-off or infiltrates water systems from broad areas [94]. It occurs when rainfall, snow-melt, or irrigation water flows across land, picking up harmful substances and depositing them into rivers, lakes, oceans, or groundwater. This form of pollution is challenging to control due to its dispersed nature, making it one of the most significant threats to water quality worldwide.

The process of NPS pollution begins when water moves over surfaces such as agricultural fields, urban areas, construction sites, and forests [79]. In agricultural settings, for instance, the over-application of fertilisers and pesticides introduces nitrogen, phosphorus, and chemical residues into water systems, as shown in Figure 2.1. These chemicals are carried by run-off during rainstorms or irrigation, eventually reaching nearby water bodies. Urban areas contribute significantly as well, with rainwater flowing across roads, rooftops, and pavements, collecting oil, grease, heavy metals, and debris before entering drainage systems that often discharge untreated water into rivers or oceans [94]. Similarly, in rural areas, animal waste and eroded soil from farms may seep into streams, contaminating them with pathogens and sediments.

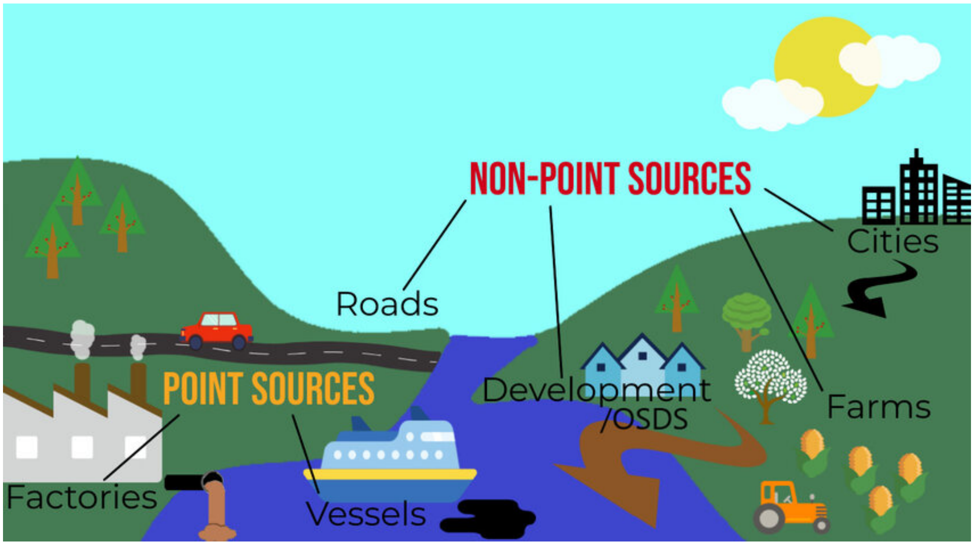

Figure 2.1: Illustration of Non-point & Point sources. Retrieved from [92]

Human activities such as deforestation, construction, and mining also exacerbate NPS pollution. When vegetation is removed, the stability of the soil diminishes, allowing sediments to be washed into nearby water bodies [128]. These sediments cloud the water, reducing light penetration and disrupting aquatic ecosystems. In addition, atmospheric deposition contributes to the problem, as airborne pollutants, such as Nitrogen Oxides (NO_x) and sulphur compounds from industrial emissions, settle onto the land and are later washed into water systems by rainfall. Recreational activities like boating and fishing can introduce pollutants such as oil, fuel, and litter into water bodies, further compounding the issue [82].

The consequences of NPS pollution are severe and far-reaching. Nutrient run-off from fertilisers and animal waste leads to the overgrowth of algae, a process known as eutrophication. This disrupts aquatic ecosystems by depleting Oxygen (O_2) levels in the water, resulting in dead zones where marine life cannot survive. A well-known example of this is the Gulf of Mexico's hypoxic zone, which forms annually due to agricultural run-off from the Mississippi River Basin [110]. Sediments from construction and deforestation cloud the water, hinder photosynthesis in aquatic plants, and clog the gills of fish. Toxic pollutants such as pesticides and heavy metals accumulate in marine organisms, entering the food chain and posing risks to both wildlife and humans. Furthermore, in agricultural regions, Nitrates (NO_3) from fertilisers frequently seep into groundwater, contaminating drinking water supplies and posing serious health risks, especially to infants [19].

Addressing NPS pollution requires a combination of preventive measures and long-term strategies. Implementing effective Best Management Practices (BMPs) is crucial for mitigating NPS pollution. These practices vary depending on the land use and the specific pollutants of concern:

Cover crops : Planting cover crops during fallow periods helps reduce soil erosion and nutrient leaching;

No-till farming : Minimising soil disturbance reduces erosion and improves water infiltration;

Riparian buffers : Planting trees and shrubs along waterways filters run-off and prevents pollutants from entering water bodies;

Nutrient management : Implementing strategies to optimise fertiliser application rates and timing reduces nutrient run-off;

Green roofs : Vegetated rooftops absorb rainwater and reduce run-off;

Permeable pavements : Allowing water to infiltrate the ground reduces run-off and recharges groundwater;

Detention ponds : These structures temporarily store stormwater run-off, allowing pollutants to settle out before the water is released;

Rain gardens : These shallow depressions planted with native vegetation capture and filter stormwater run-off;

Erosion control measures : Using silt fences, erosion blankets, and other measures to prevent soil erosion during construction;

Sediment fences : These barriers trap sediment run-off, preventing it from entering water bodies.

Therefore, NPS pollution requires a combination of preventive measures and long-term strategies. Changes in land-use practices, such as adopting conservation agriculture or creating buffer zones along waterways, can significantly reduce runoff [128]. Green infrastructure in urban areas, like rain gardens and permeable pavements, helps manage stormwater by allowing it to filter into the ground rather than flowing untreated into waterways. Regulatory measures can also play a role, such as limiting the use of fertilisers and pesticides, especially near water bodies. Public awareness campaigns that educate communities about the impacts of their actions on water quality are equally vital. Lastly, advancements in monitoring and modelling tools allow researchers to better understand and mitigate the effects of NPS pollution [9].

The second primary source of pollution is called the PS Pollution refers to water contamination that originates from a single, identifiable source, such as a pipe, ditch, or conduit [82]. Unlike NPS pollution, which is diffuse and scattered across large areas, PS pollution can be traced directly to its origin, making it easier to identify, monitor, and regulate, as shown in Figure 2.1. However, despite its localised nature, PS pollution can have devastating effects on aquatic ecosystems and human health if not adequately controlled [56].

This type of pollution typically occurs when pollutants are discharged directly into a water body from industrial facilities, Wastewater Treatment Plants (WWTPs), or other concentrated sources [28]. For instance, factories often release untreated or inadequately treated wastewater containing harmful chemicals, heavy metals, and organic waste directly into rivers or lakes. Similarly, municipal wastewater systems, when overwhelmed by heavy rainfall or poor maintenance, can release untreated sewage into nearby water bodies. Thermal pollution, where factories or power plants discharge heated water into rivers or lakes, is another form of PS pollution. This sudden increase in Water Temperature (WT) disrupts aquatic ecosystems, reducing O_2 levels and endangering species that thrive in cooler environments [96].

Agricultural operations can also contribute to PS pollution. In some cases, animal feeding operations, such as large-scale livestock farms, generate concentrated waste that is stored in lakes or tanks [39]. If these containment systems leak or overflow, they can release high levels of nutrients, pathogens, and other pollutants directly into nearby streams or groundwater. Mining operations are another significant contributor, as runoff from mine sites can introduce heavy metals such as mercury, arsenic, and cadmium into surrounding water systems [94].

Urban areas often exacerbate PS pollution through stormwater systems that discharge directly into water bodies [81]. These systems collect run-off from roads, car parks, and industrial zones, channelling pollutants such as oil, grease, and toxic chemicals into nearby rivers or oceans. Leaking underground storage tanks at gas stations or industrial sites can also act as point sources, contaminating both surface and groundwater with petroleum products [19].

The environmental consequences of PS pollution are profound. Toxic substances discharged into water bodies can poison aquatic life, disrupt reproductive cycles, and accumulate in the food chain, affecting both wildlife and humans [136]. Nutrient-rich discharges, such as those from Sewage Treatment Plants (STPs), can trigger algal blooms that deplete O_2 levels in the water, causing fish kills and the destruction of aquatic habitats. Heavy metals and persistent organic pollutants, which do not readily degrade, can persist in the environment for decades, posing long-term risks to ecosystems and human health. For example, mercury pollution from industrial sources has contaminated fish in rivers and lakes, making them unsafe for consumption [82].

PS pollution is also a significant public health concern. Contaminants such as pathogens, heavy metals, and industrial chemicals can seep into drinking water supplies, causing illnesses and chronic health conditions [112]. Communities near industrial facilities or sewage outfalls are often disproportionately affected, highlighting the need for equitable enforcement of pollution regulations [124].

One of the key advantages of addressing PS pollution is that it is relatively straightforward to regulate compared to non-point sources. Governments and environmental agencies can enforce laws and issue permits that set limits on the types and amounts of pollutants a facility can discharge [140]. Technologies such as wastewater treatment systems, scrubbers, and containment measures have been developed to minimise the impact of PS pollution. For instance, many countries have implemented policies requiring factories to treat their effluents before releasing them into water bodies, significantly reducing the levels of toxic substances in the environment [48].

Governments and environmental agencies play a crucial role in controlling PS pollution through regulations and permitting systems [140]. In the United States, the Clean Water Act (CWA) establishes the basic structure for regulating discharges of pollutants into the waters of the United States. The CWA requires industries and municipalities to obtain permits for any discharges, setting limits on the types and amounts of pollutants that can be released. These permits often require the implementation of specific treatment technologies and regular monitoring to ensure compliance [118]. Similar regulatory frameworks exist in other countries.

However, enforcement of these regulations can be challenging. Limited resources, political pressure, and complex legal processes can hinder effective enforcement. Monitoring is essential to ensure compliance with permit limits, but it can be expensive and time-consuming. Furthermore, some industries may try to circumvent regulations or fail to report violations. Strengthening enforcement mechanisms, increasing funding for monitoring, and promoting transparency are crucial for effective control of PS pollution [48].

WWTP play a vital role in removing pollutants from municipal and industrial wastewater before it is discharged into water bodies, as illustrated in Figure 2.2 [28]. Treatment typically involves several levels:

Primary Treatment : This involves physical processes like screening and settling to remove large solids and debris [28];

Secondary Treatment : It uses biological processes to remove organic matter, typically by using microorganisms to break down the pollutants [28];

Tertiary Treatment : The last step involves advanced treatment processes to remove specific pollutants, such as nutrients (nitrogen and phosphorus), heavy metals, or pathogens. Technologies used in tertiary treatment include advanced filtration, reverse osmosis, and disinfection [28].

Despite these advancements, PS pollution remains a persistent issue in many parts of the world, particularly in developing countries where environmental regulations are weak or poorly enforced [44]. The rapid growth of industrial activities and urbanisation has increased the pressure on existing infrastructure, leading to frequent pollution incidents. Climate change further complicates the problem, as extreme

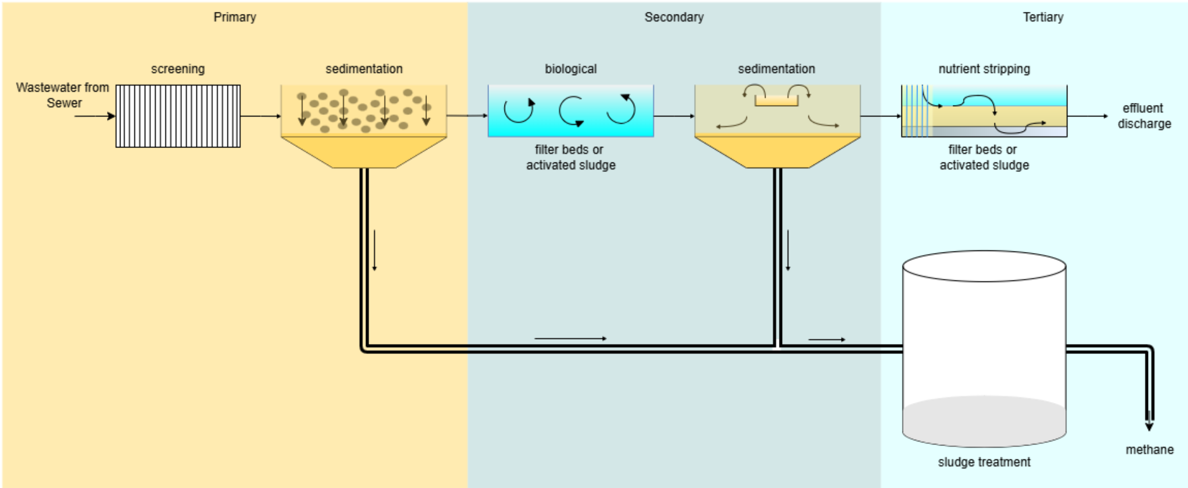

Figure 2.2: Wastewater Treatment Plant. Adapted from [28]

weather events such as floods and hurricanes can overwhelm wastewater systems, leading to untreated sewage and industrial waste entering water bodies [141].

Efforts to combat PS pollution must focus on strengthening regulations, investing in infrastructure, and promoting cleaner production methods [28]. Public awareness campaigns can also play a critical role in encouraging industries and communities to adopt sustainable practices. By addressing PS pollution effectively, it is possible to protect water quality, safeguard ecosystems, and ensure access to clean water for future generations [58].

The third-largest source of contamination is Transboundary water pollution, which occurs when contamination crosses the political or geographic boundaries of one country to affect another [13]. This type of pollution highlights the interconnected nature of water systems, where rivers, lakes, and aquifers span multiple countries, and actions upstream can have significant consequences downstream. Transboundary water pollution is often the result of activities such as industrial discharges, agricultural runoff, mining operations, or wastewater treatment failures in one region, leading to the introduction of pollutants that travel through shared water systems [43].

One of the primary ways this happens is through rivers that flow across national borders. For example, the Tagus River, which originates in Spain and flows into Portugal, has experienced significant pollution issues over the years [99]. Industrial and agricultural activities upstream in Spain have released contaminants, including heavy metals, pesticides, and nutrient-rich run-off, into the river. By the time the river reaches Portugal, these pollutants can have detrimental effects on water quality, aquatic ecosystems, and human populations relying on the water for drinking, irrigation, and recreation. During periods of drought, when water flow decreases, pollutant concentrations increase, exacerbating the problem [99].

It also occurs through shared groundwater resources or aquifers [13]. When pollutants such as NO_3 from fertilisers, heavy metals, or chemicals seep into the groundwater in one country, they can travel through underground water systems and contaminate resources in neighbouring regions. This often leads to long-term environmental challenges, as groundwater contamination is difficult to detect and remediate.

Another example is the Douro River, which also flows from Spain into Portugal. Similar to the Tagus, the Douro has been affected by upstream activities, such as industrial effluents and untreated wastewater [114]. These discharges introduce toxic substances into the river, affecting downstream ecosystems and posing risks to public health. Fish populations in Portuguese sections of the river have been adversely affected, disrupting local fishing economies and biodiversity [114].

In addition to direct discharges, transboundary water pollution can result from diffuse sources, such as agricultural run-off containing excess fertilisers and pesticides [128]. These pollutants enter rivers and streams through rainfall or irrigation systems and are carried across borders, leading to eutrophication in downstream lakes and reservoirs. This process stimulates excessive algal growth, depleting O_2 levels and causing harm to aquatic life [86].

The geopolitical nature of shared water resources further complicates the issue. Upstream countries often have significant control over the quantity and quality of water that flows downstream, which can lead to conflicts over resource use and pollution levels [13]. In some cases, upstream countries may prioritise their own economic activities, such as intensive agriculture or industrial production, without fully considering the environmental and social impacts on downstream nations [85].

Efforts to address transboundary water pollution require international cooperation and agreements. In Europe, the Albufeira Convention serves as an example of such an effort [89]. Signed by Portugal and Spain in 1998, this treaty establishes guidelines for the sustainable management of shared river basins, aiming to balance water use, protect ecosystems, and ensure equitable access to clean water for both countries. However, tensions can still arise, particularly during droughts or when upstream pollution incidents occur, as the effectiveness of such agreements often depends on mutual trust and compliance [13].

The impact can be visually seen in Figure 2.3, where, for example, upstream of a river, there might be rain besides it, and sediments from the ground can leak into the water body. Agriculture, together with livestock of animals, which are typically near a water body, particularly a river, can leak contaminants into the river [79]. Especially in an agricultural context, they usually use pesticides and other nutrient-based fertilisers, which are applied to soils or vegetation and can enter the freshwater environment. The impact is the creation of algal blooms, which are specific areas in the water body with low or no O_2. Industries and households, which are typically built out of different materials such as copper or zinc, degrade over time. During heavy rainstorms, these metals can be carried by the water and deposited into the different waterways [131]. The work of other industries, which produce various products, generates significant waste, much of which runs off into freshwater systems. At the end of many streams, swimming, fishing, or drinking is not allowed due to heavy contamination. These can make the population ill and cause death from pathogens. This has a profound cultural impact, as humans cannot enjoy or practice their day-to-day lives [81].

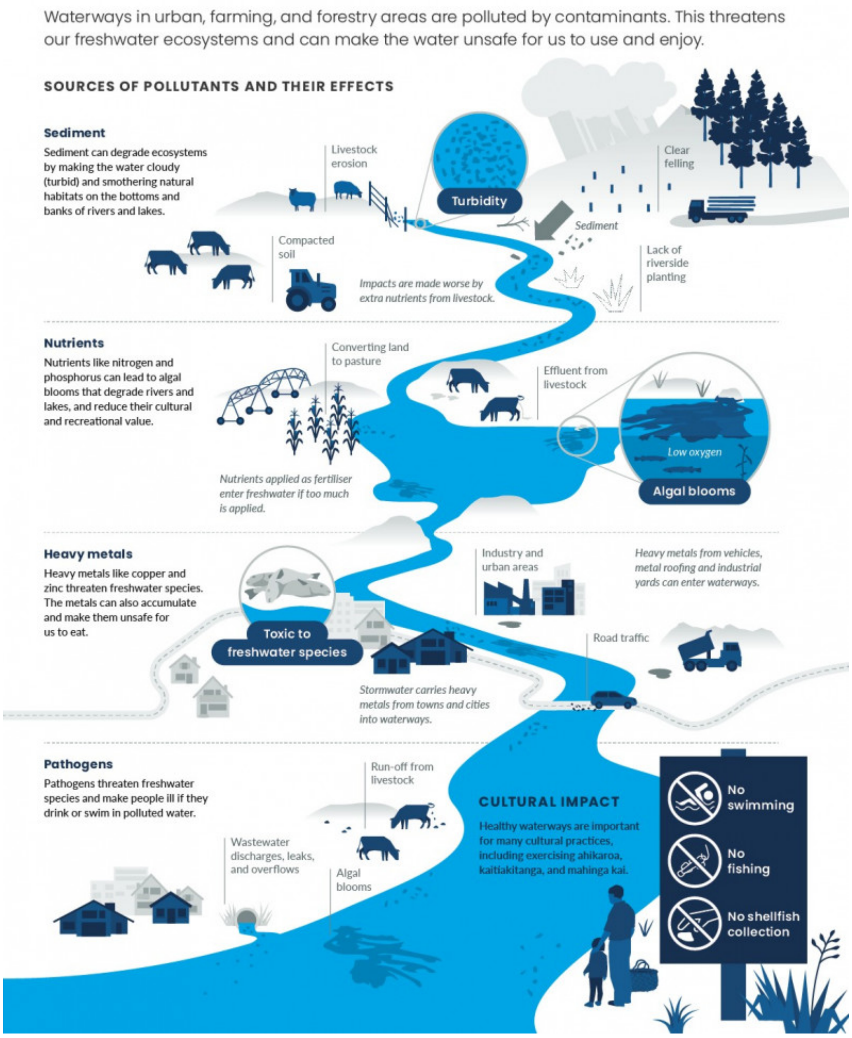

Figure 2.3: Impact of Water Pollution. Retrieved from [100]

2.2.1 Water Pollutants

Water pollution originates from substances or nutrients that, when introduced into water bodies, alter their natural composition, degrade their quality, and harm ecosystems, aquatic life, and human health. These pollutants can originate from natural sources or human activities and enter water systems through various pathways, such as direct discharge, run-off, atmospheric deposition, or subsurface infiltration [94]. The introduction of these substances into water disrupts the ecological balance, reduces the water's usability, and poses risks to organisms that depend on it.

These pollutants can be classified into different categories, such as agricultural, atmospheric, pathogens, chemical and physical [19].

2.2.1.1 Agricultural

The first category is a significant contributor to water pollution, primarily originating from farming activities and the use of agrochemicals [128]. Fertilisers, pesticides, herbicides, and animal waste introduce harmful substances into water systems, particularly through run-off during rainfall or irrigation. Fertilisers, which are rich in nitrogen and phosphorus, are used to enhance crop productivity but often lead to nutrient overloading in nearby water bodies [94]. This excessive nutrient input can cause eutrophication, a process that stimulates algal blooms, depletes O_2 levels, and disrupts aquatic ecosystems. Pesticides and herbicides, designed to protect crops from pests and weeds, contain chemical compounds that can be toxic to non-target species, including marine life and humans. These substances often persist in the environment, accumulate in sediments, and enter the food chain. Phosphates, which are commonly found in fertilisers and detergents, further exacerbate nutrient pollution and degrade water quality. Animal farming operations, including concentrated animal feeding operations, generate organic waste, pathogens, and nutrients that, when washed into rivers and lakes, cause significant water quality problems. Agricultural pollutants are especially problematic in regions with intensive farming practices, where land management and water protection are insufficiently regulated [39].

An example is Lake Erie, one of North America's Great Lakes, which has been significantly impacted by agricultural pollutants, particularly from phosphorus-rich fertilisers used in farming [141]. During the 1960s and 1970s, agrarian run-off containing these nutrients flowed into the lake, causing severe eutrophication. The resulting algal blooms led to O_2 depletion, widespread fish kills, and disruptions in the aquatic ecosystem. This issue resurfaced in 2014 when harmful algal blooms caused by microcystin, a toxic cyanobacterium, contaminated Toledo, Ohio's water supply, leaving half a million residents without safe drinking water. The case highlights the long-term environmental consequences of intensive farming practices and the importance of regulating agricultural run-off [48].

2.2.1.2 Atmospheric

Atmospheric pollutants are substances released into the atmosphere that eventually settle into water bodies through processes such as precipitation, dry deposition, or gas exchange [94]. CO_2s and nitrogen compounds are among the most significant contributors to water pollution via atmospheric deposition. Elevated levels of CO_2s in the atmosphere dissolve in water, forming carbonic acid and contributing to ocean acidification, which disrupts marine ecosystems by affecting organisms like shellfish and corals that rely on calcium carbonate to build their shells and skeletons. Nitrogen compounds, such as NO_x and Ammonia (NH_3), are released into the air from industrial activities, vehicle emissions, and agricultural practices [79]. These compounds are carried by wind and eventually deposited into water bodies, where they contribute to nutrient pollution. The resulting nitrogen enrichment promotes algal blooms and leads to hypoxic conditions harmful to aquatic life. Additionally, atmospheric pollutants like Sulphur Dioxide (SO_2), released from fossil fuel combustion, combine with water vapour to form acid rain, which can lower the Potential of Hydrogen (pH) of lakes and rivers, harming fish and other aquatic organisms. Atmospheric deposition is a global issue, as pollutants can travel long distances before settling into water systems, causing transboundary water pollution [81].

This has happened with acid rain, primarily caused by atmospheric pollutants such as SO_2 and NO_x, which became a significant issue in the northeastern United States during the mid-20th century [81]. These pollutants were released into the atmosphere from industrial facilities and power plants burning fossil fuels, and they combined with water vapour to form acidic precipitation. This acid rain damaged forests, acidified lakes, and killed aquatic life across the region. For example, New York's Adirondack Lakes became so acidic that fish populations were decimated. The phenomenon was mitigated after the Clean Air Act amendments of 1990 introduced regulations on SO_2 and NO_x emissions, demonstrating the impact of atmospheric pollutants on water systems and the benefits of environmental policies [66].

2.2.1.3 Pathogens

Pathogens are microorganisms, including bacteria, viruses, and protozoa, that contaminate water and pose serious risks to human and animal health. These biological pollutants often originate from untreated sewage, agricultural runoff, or improperly managed livestock waste. Common waterborne pathogens include Salmonella, Escherichia Coli (E. coli), Cryptosporidium, and Vibrio cholerae, which cause diseases ranging from gastrointestinal illnesses to severe outbreaks of cholera and typhoid fever [112]. Pathogens enter water systems through various pathways, including the direct discharge of untreated or inadequately treated wastewater, runoff from agricultural fields containing animal waste, and stormwater overflows. Once present in water, these microorganisms can survive for extended periods, depending on environmental conditions. The health impacts of pathogen-contaminated water are particularly severe in regions with limited access to clean water and sanitation facilities, where outbreaks of waterborne diseases can lead to high morbidity and mortality rates. In addition to human health risks, pathogens in water can harm aquatic life, causing diseases in fish and other organisms, and disrupt ecosystems [124].

The Ganges River, one of the most significant water bodies in India, is heavily polluted with pathogens, primarily due to the discharge of untreated sewage, industrial effluents, and agricultural run-off [131]. Pathogens such as E. coli and cholera-causing bacteria are prevalent in the river, posing significant health risks to the millions of people who rely on it for drinking water, bathing, and agriculture. Outbreaks of waterborne diseases are common in areas along the Ganges, particularly during religious festivals, when millions of people gather to perform rituals in the river. Efforts such as the "Namami Gange" program have been initiated to clean and restore the river. Still, challenges remain due to the scale of pollution and the lack of adequate sewage treatment infrastructure [131].

2.2.1.4 Chemical

Chemical pollutants encompass a wide variety of substances that contaminate water, altering its chemical composition and causing ecological and health issues [82]. These pollutants include dissolved salts, acids, heavy metals, fluoride, chlorides, NO_3, phosphates, ammonium, organic compounds, and agrochemicals such as pesticides and herbicides. Dissolved salts, often originating from irrigation run-off, industrial discharges, or natural mineral deposits, can increase water salinity and make it unsuitable for drinking or agricultural use. Acids, typically from industrial effluents or acid rain, lower the PH of water, harming aquatic organisms and corroding infrastructure. Heavy metals such as lead, mercury, and cadmium are toxic even in small amounts, bioaccumulate in marine organisms, and pose severe health risks when they enter the human food chain. Fluoride and chlorides, while beneficial in small amounts, can be harmful at elevated concentrations, leading to toxicity and water quality degradation. NO_x and phosphates, primarily from agricultural run-off and wastewater, contribute to nutrient pollution, promoting eutrophication and harming aquatic ecosystems. Organic compounds, including petroleum products, solvents, and plastics, persist in the environment and pose long-term risks to ecosystems and human health. Pesticides and herbicides, widely used in agriculture, contain active ingredients that can be toxic to non-target organisms, disrupt ecosystems, and contaminate groundwater [128]. Ammonium contributes to the nitrogen cycle and exacerbates nutrient imbalances. Excessive ammonium in aquatic systems promotes algal blooms and increases water toxicity. A direct measure of it is Chlorophyll-a (Chl-a), which is a key indicator of algal biomass and a direct measure of nutrient pollution [138]. High Chl-a levels signal nutrient-rich environments that can lead to hypoxic conditions, disrupting aquatic ecosystems and threatening water quality for human use. The complexity and diversity of chemical pollutants make their management challenging, requiring stringent regulation and advanced treatment technologies [48].

In Japan, the Minamata Bay became infamous for one of history's most devastating examples of chemical water pollution [147]. Between the 1930s and 1960s, a chemical plant owned by Chisso Corporation released methylmercury into the bay as part of its industrial waste. This toxic compound bioaccumulated in fish and shellfish, which were a dietary staple for residents. Consuming contaminated seafood caused Minamata Disease, a neurological syndrome that led to severe physical impairments, developmental delays in children, and even death. The disaster highlighted the dangers of industrial chemical discharge into water systems and led to global awareness of the need for stricter industrial waste management practices [147].

2.2.1.5 Physical

Physical pollutants are substances that alter the physical properties of water, affecting its clarity, temperature, and flow dynamics. These pollutants include both organic and inorganic materials suspended in water, such as sediments, silt, and debris [94]. Suspended sediments, often resulting from soil erosion, deforestation, or construction activities, increase water turbidity, reducing sunlight penetration and disrupting photosynthesis in aquatic plants. High turbidity can also clog fish gills, smother benthic habitats, and interfere with the natural flow of water bodies. Organic materials, such as plant debris and decaying matter, contribute to O_2 depletion as they decompose, negatively affecting aquatic ecosystems. Temperature changes, often caused by thermal pollution from industrial discharges or deforestation, can alter water quality by reducing O_2 solubility and stressing temperature-sensitive aquatic species [19]. Physical pollutants also include debris like plastic particles, which, while initially a physical issue, can fragment into microplastics and pose chemical and biological risks over time. Managing physical pollutants is critical for maintaining the ecological balance of water systems and ensuring their suitability for human and environmental use [69].

The Great Pacific Garbage Patch, located in the North Pacific Ocean, is a massive accumulation of physical pollutants, primarily plastics [42]. This gyre, formed by ocean currents, has trapped millions of tons of debris, including microplastics, fishing gear, and consumer waste. The patch's formation is a result of decades of plastic pollution entering the oceans from rivers, coastal areas, and marine activities. Debris affects aquatic life, as animals often mistake plastic for food or become entangled in larger items. The Great Pacific Garbage Patch has become a symbol of the global plastic pollution crisis, prompting international efforts to reduce single-use plastics and improve waste management [42].

2.2.2 Impact of Pollution on Different Water Bodies

The environmental impact of water pollution varies significantly depending on the type of water body affected. Water bodies can be broadly categorised as still water (lentic), including lakes and ponds, and flowing water (lotic), encompassing rivers, streams, oceans, and seas. These distinct hydrological characteristics profoundly influence how pollutants behave and the ultimate effects they have on ecosystems.

Still water bodies, such as lakes and ponds, are characterised by limited water flow, making them particularly vulnerable to the accumulation of pollutants. These water bodies can also be categorised as Lentic Systems. Their relative lack of mixing can lead to stratification, a phenomenon in which distinct layers form within the water column based on temperature, density, or chemical composition [16]. Figure 2.4 helps to visualise such phenomena. Stratification can result in O_2 depletion (hypoxia or anoxia) in deeper layers (hypolimnion), as the lack of vertical mixing prevents the replenishment of O_2 from the atmosphere. This O_2 depletion can have devastating consequences for aquatic life, as many organisms require dissolved O_2 to survive [37].

Lakes also tend to act as sinks for pollutants. Sediments, nutrients, and other contaminants can settle and accumulate over time, leading to long-term contamination and gradual deterioration of water quality [86]. Nutrient pollution, primarily from agricultural run-off containing nitrogen and phosphorus, is a significant concern in lakes. Excess nutrients can trigger eutrophication, a process characterised by excessive algal growth (algal blooms) [121]. These blooms block sunlight, hindering the growth of submerged aquatic plants. When the algae die and decompose, they further deplete O_2 levels, exacerbating hypoxia and creating "dead zones" where marine life cannot survive. The long residence time of water in lakes also means that pollutants can remain in the system for extended periods, making remediation efforts more challenging [86].

The other major water body systems are flowing water bodies, such as rivers and streams, which are more dynamic and are also part of the Lotic Systems. The constant flow helps to disperse pollutants, potentially mitigating localised impacts [37]. Rivers can transport contaminants downstream, affecting ecosystems far from the original pollution source. This downstream transport can have transboundary implications, affecting water quality and ecosystems in other regions or even countries [13]. However, the

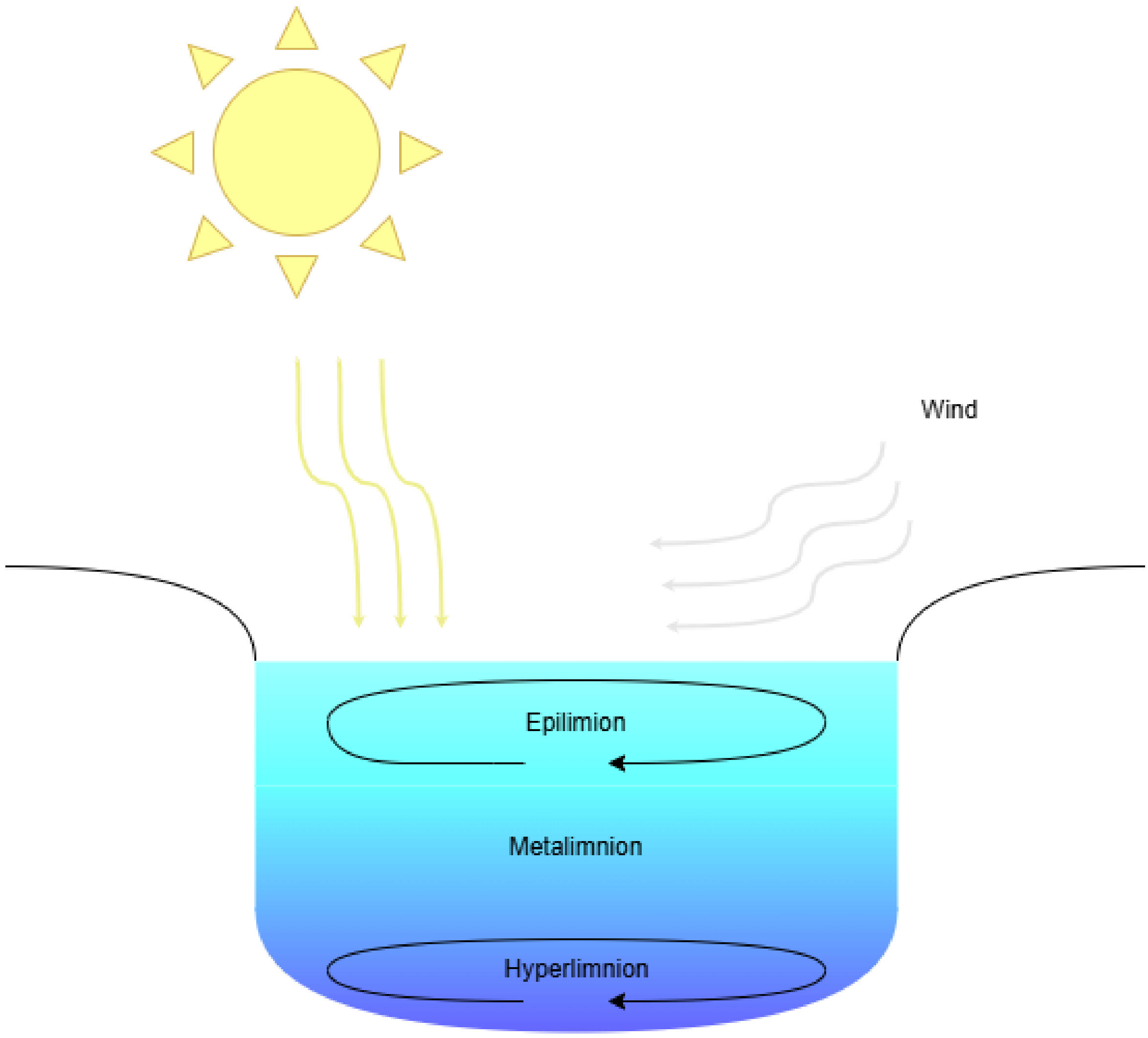

Figure 2.4: Lake stratification. Adapted from [52]

continuous flow also facilitates faster dilution and aeration, which can help to break down some pollutants and replenish dissolved O_2.

Oceans and seas, while vast, are not immune to pollution. They face unique challenges, including oil spills, plastic accumulation (particularly in gyres), and the concentration of pollutants in coastal areas where water circulation may be restricted [116]. Plastic pollution is a significant concern, as plastics can fragment into microplastics, which can be ingested by marine organisms and enter the food chain. Ocean acidification, caused by the absorption of atmospheric CO_2, is another growing problem that impacts marine ecosystems, particularly shell-forming organisms. Coastal areas, which are often highly productive and support diverse ecosystems, are particularly vulnerable to pollution from both land-based and marine sources [37].

Additionally, it is crucial to recognise that these different types of water bodies are interconnected. Pollution in rivers can eventually flow into oceans and seas, and contaminated groundwater can seep into lakes and streams. Therefore, a holistic approach to water pollution management is essential, considering the entire watershed and the interconnectedness of all water bodies [13].

2.2.3 National Events

A more national event was a study conducted on a Portuguese river, particularly the Ave river, which goes across the whole country from Spain to the Atlantic Ocean in northern Portugal, has a history of water quality degradation due to industrial discharges [10]. Research assessing the river's water quality identified significant physicochemical contamination, attributed to effluents from textile and other industries that discharge directly into the river. These PS pollutants have led to the deterioration of aquatic ecosystems and raised concerns over public health and environmental safety. The Ave River's water quality is not suitable for irrigation or human consumption due to high faecal contamination, despite its apparent good ecological status, as indicated by macroinvertebrates and physicochemical parameters [10].

Another study of the Tinto River, which flows through Porto's urban area, has identified significant contamination [76]. Studies conducted in September 2022 revealed elevated levels of total aerobic microorganisms, coliforms, and Enterococcus, indicating faecal contamination. The presence of these pathogens suggests that untreated or inadequately treated wastewater is being directly discharged into the river. There was a significant contamination disparity between the Tinto River source and the other studied locations, with notably higher contamination levels compared to the source, a clear example of PS pollution. Although a project to stop the release of effluents from two WWTPs into the Tinto River was completed in 2019, the river still exhibits significant contamination [76].

2.3 Remote Sensing

RS refers to the science and technology of acquiring information about the Earth's surface or atmosphere without being in direct contact with it. It relies on sensors to detect and measure reflected or emitted radiation, providing insights into a range of natural and human-made phenomena. The field involves a combination of satellite systems, sensors, and sophisticated data processing techniques, making it indispensable for studying the Earth on a global scale [35].

It enables data collection over extensive areas at varying spatial and temporal resolutions. In addition to these, radiometric resolution is a key characteristic that describes a sensor's sensitivity to detect subtle differences in energy intensity [119]. Higher radiometric resolution allows the identification of more subtle variations in surface reflectance or emitted radiation, which is crucial for detailed environmental monitoring. Sensors are classified into active and passive types. Active sensors, like radar, generate their own energy to illuminate the target and measure the reflected signal. Passive sensors, such as optical cameras, depend on external energy sources, such as sunlight, to detect reflected or emitted radiation. This versatility has made RS a fundamental technique in fields such as climate science, agriculture, urban development, and water quality management [35].

Before analysis, raw satellite data typically undergoes preprocessing steps, such as atmospheric correction, radiometric calibration, geometric correction, and cloud masking. For example, Sentinel-2 data can be processed with the Sen2Cor [111] algorithm to correct for atmospheric effects, such as aerosols and water vapour. These steps are essential to ensure that derived indices and measurements are physically consistent and comparable across time and space.

Despite its strengths, remote sensing has some limitations. Optical imagery is often hindered by cloud cover and atmospheric distortions, which restrict data availability in certain regions [119]. Radar systems overcome cloud interference but can be more complex to interpret. [107] Other challenges include spectral confusion (different materials with similar reflectance signatures), the need for ground truth data for validation, and the computational intensity of processing large datasets. Recognising these limitations is essential for correctly interpreting results and combining satellite data with in-situ observations [35].

2.3.1 Copernicus Programme

The Copernicus Programme, launched by the EU in collaboration with the European Space Agency (ESA), is one of the most advanced and comprehensive Earth observation systems in the world [138]. Operational since 2014, it aims to monitor Earth's environment and support European policymaking across domains such as climate change, disaster management, agriculture, and urban planning. The programme's primary goal is to provide high-quality, timely, and freely accessible data to address pressing environmental and societal challenges on a global scale [71].

It evolved from the Global Monitoring for Environment and Security (GMES) initiative, which was conceived in the late 1990s as a response to the growing need for Earth observation data to support environmental management and security [138]. While GMES initially focused on research-based observations, it later transitioned into a fully operational service with the launch of the first Sentinel satellite in 2014 [107]. This shift marked the beginning of the Copernicus programme's expansive role in real-time data collection and dissemination.

Copernicus benefits from an extensive infrastructure that includes a constellation of satellites known as the Sentinel series, each designed to monitor specific aspects of Earth's surface and atmosphere [71]. These satellites cover a range of observation needs, from land and ocean monitoring to atmospheric analysis and emergency response [138]. In addition to the Sentinel satellites, the Copernicus programme is supported by a network of ground-based stations, which provide critical calibration data and support the collection of supplementary information. Data analysis systems further enhance the programme's ability to process and interpret vast amounts of satellite data in real time.

The data that comes from this programme has proven instrumental in a wide range of applications. For example, its contributions to monitoring deforestation and land use changes have helped researchers understand the global impacts of climate change and biodiversity loss [71]. In water management, Copernicus data supports the monitoring of water quality, including key parameters such as turbidity and chlorophyll levels, which are essential for assessing ecosystem health and managing freshwater resources [138]. Furthermore, the Copernicus programme is crucial for predicting and managing natural disasters like floods, wildfires, and coastal erosion. The programme's emphasis on open-access data has encouraged collaboration among researchers, industry professionals, and policymakers, fostering innovation in environmental monitoring, resource management, and sustainable development. By providing timely and accurate data, Copernicus enables more informed decision-making, contributing to the EU's broader goals of environmental sustainability and climate resilience [119].

2.3.2 USGS Programme

The United States Geological Survey (USGS) manages the Landsat programme, one of the longest-running and most influential Earth observation initiatives globally. Launched in 1972 in collaboration with National Aeronautics and Space Administration (NASA), the Landsat programme has continuously provided the world with a comprehensive, high-resolution dataset of Earth's land surface [142]. With over four decades of data, the Landsat programme is instrumental in understanding the dynamics of land use, ecosystems, and urbanisation on a global scale.

The Landsat programme has had a profound impact on scientific research and environmental monitoring. The programme's data has been used extensively to track long-term changes in land cover and land use. For example, researchers have used Landsat data to monitor urban expansion, detect deforestation, and study the effects of climate change on ecosystems [143]. Landsat's historical archive has also enabled trend analyses spanning decades, allowing scientists to gain insights into global land surface dynamics.

Since 2008, the USGS has made Landsat data freely available to the public, a transformative move that has revolutionised environmental research. This open-access policy has expanded the use of Landsat data beyond the scientific community to include government agencies, industry professionals, Nongovernamental Organizations (NGOs), and even individuals [142]. Free access to reliable, consistent data has empowered researchers to conduct in-depth studies on topics such as land degradation, agricultural productivity, and the monitoring of natural hazards, including wildfires and droughts [90].

In addition to its environmental applications, Landsat data has been invaluable for urban planning and disaster management [149]. By providing long-term, accurate data, the Landsat programme has supported the development of tools for monitoring urban growth, assessing the impacts of natural disasters, and planning for sustainable land use [90]. The enduring legacy of the Landsat programme, with its extensive archive and continuous data streams, ensures its continued relevance in global Earth observation efforts. Landsat remains a cornerstone of environmental monitoring, helping policymakers and researchers make informed decisions in response to the challenges posed by rapid environmental and societal changes [143].

2.3.3 Satellite Imagery

Satellite imagery refers to visual or spectral data captured by sensors onboard satellites orbiting the Earth [35]. These images provide vital information about the planet's surface, atmosphere, and oceans. A typical satellite consists of a payload (the sensors and instruments), a platform (the satellite structure), and communication systems to relay data to ground stations [20].

Satellite imagery is characterised by three main resolutions:

Spatial resolution represents the smallest object that can be distinguished in an image, often expressed as pixel size. High spatial resolution imagery is essential for detailed studies, such as urban mapping or infrastructure monitoring, while lower resolutions are more suitable for broader regional or global assessments [35].

Spectral resolution refers to the capability of a sensor to capture data at specific wavelengths across the electromagnetic spectrum. This enables the identification of various materials, vegetation health, or water quality through unique spectral signatures. For instance, multispectral sensors capture data in a limited number of broad bands, while hyperspectral sensors provide finer details by dividing the spectrum into many narrow bands [20].

Temporal resolution indicates the frequency at which a satellite revisits and images a particular area. High temporal resolution is crucial for monitoring dynamic phenomena, such as deforestation, crop growth, or disaster progression, as it allows for regular updates over time [125].

These characteristics make satellite imagery an indispensable resource for observing and understanding the Earth's complex systems. By combining data from different resolutions, scientists and researchers can extract valuable information for environmental monitoring, disaster response, resource management, and policymaking [20]. Advances in satellite technology, including the development of constellations like Sentinel-2 and Landsat, have further enhanced the availability, accuracy, and accessibility of satellite imagery, driving innovation across various fields of study [138, 142].

2.3.3.1 Sentinel

Part of the Copernicus Programme is the Sentinel constellation of satellites; there are 7 in total. Each Sentinel is considered a mission [71].

Sentinel-1 is equipped with Synthetic Aperture Radar (SAR) technology, which provides all-weather, day-and-night observations, beneficial for flood mapping, sea ice monitoring, and agricultural assessments. It consists of two satellites, the first launched in April 2014 and the second in April 2016. The second satellite ended in 2022, with plans to launch another one as soon as possible [107]. This satellite is portrayed as an active radar, meaning it does not need sunlight to use its wavelengths to produce an image. It instead directly shoots radar, at a wavelength of 56mm, offering the possibility of penetrating clouds.

Sentinel-2 focuses on optical imagery for land and vegetation monitoring; it offers data critical for soil and water analysis. It can also deliver information for emergency services. This mission consists of two satellites: the first, Sentinel-2A, was launched in June 2015, and the second, Sentinel-2B, was launched in March 2017 [125].

Sentinel-3 is designed for sea and land surface monitoring, measuring variables like sea-surface temperature, ocean colour, and vegetation indices. The mission includes support for ocean forecasting systems and environmental and climate monitoring. It consists, again, of two satellites, launched in February 2016 and April 2018 [23].

Sentinel-4's geostationary satellites monitor atmospheric composition, focusing on air quality and trace gases over Europe. It is expected to be launched in 2025 as a payload on Meteosat Third Generation Sounders (MTG-S) [24].

Sentinel-5 Precursor is dedicated to global atmospheric monitoring, tracking air pollutants and greenhouse gases, aiding climate studies and public health initiatives. It was taken into orbit in October 2017 [113].

Sentinel-5 will be onboard Meteorological Operational Satellite - Second Generation (MetOp-SG) satellites, providing enhanced atmospheric composition data for climate monitoring and public health applications [65].

Sentinel-6's mission is to measure changes in global sea level, which are directly linked to climate change, making it a key indicator. It maps 95% of Earth's ice-free water surface, including ocean currents, wind speed, and wave height. The first was launched in 2020, and the second in 2025. [31].

Sentinel-2 features 13 spectral bands ranging from visible to Short-Wave Infrared (SWIR) wavelengths [30], highlighted in Figure 2.5, where the main ones are as follows:

Band 1 is also called Coastal aerosol, operates at 443 nm, is used for atmospheric correction, and is particularly useful for monitoring coastal regions and mapping shallow waters, where its shorter wavelength can penetrate water effectively.

Band 2 is also known as the blue band; when used in Red-Green-Blue (RGB), it corresponds to the blue channel at 490 nm. It is helpful for soil and vegetation discrimination, forest type mapping and identifying man-made features. It is scattered by the atmosphere, it illuminates material in shadows better than longer wavelengths, and it penetrates clear water better than other colours. It is absorbed by chlorophyll, which results in darker plants.

Band 3 corresponds to the green spectrum, therefore the green channel in an RGB image, operating at 560 nm. It gives excellent contrast between clear and turbid (muddy) water, and penetrates clear water fairly well. It helps in highlighting oil on water surfaces and vegetation. It reflects green light strongly than any other visible colour. Man-made features are still visible.

Band 4 is the use case for the red channel, meaning it uses the 665 nm spectral band. It is strongly reflected in dead foliage and helps identify vegetation types, soil types, and urban areas (cities and towns). It has limited water penetration and doesn't reflect well from live foliage with chlorophyll.

Band 5-7 is critical for plant health studies, being sensitive to chlorophyll content, also known as the red edge at a wavelength of 705-783 nm.

Band 8 enters the near infrared spectrum, where it is suitable for mapping shorelines and biomass content, as well as for detecting and analysing vegetation, working at 842 nm.

Band 8A provides more precise vegetation differentiation than Band 8, by being narrower in the Near-Infrared (NIR) space, working at a wavelength of 865 nm.

Band 9 offers the possibility to detect water vapour, where it's spectral wave is at the 945nm.

Band 10 provides a way to detect the cirrus clouds, which are the ones that are really high and thin. This band is at 1375nm.

Band 11 operates at 1610 nm, which defines the SWIR, helps measure soil and vegetation moisture content, and provides good contrast between different types of vegetation. It helps differentiate between snow and clouds. On the other hand, it has limited cloud penetration.

Band 12 is similar to the one before, with the same use case, differing only in that it operates in the 2190 nm range.

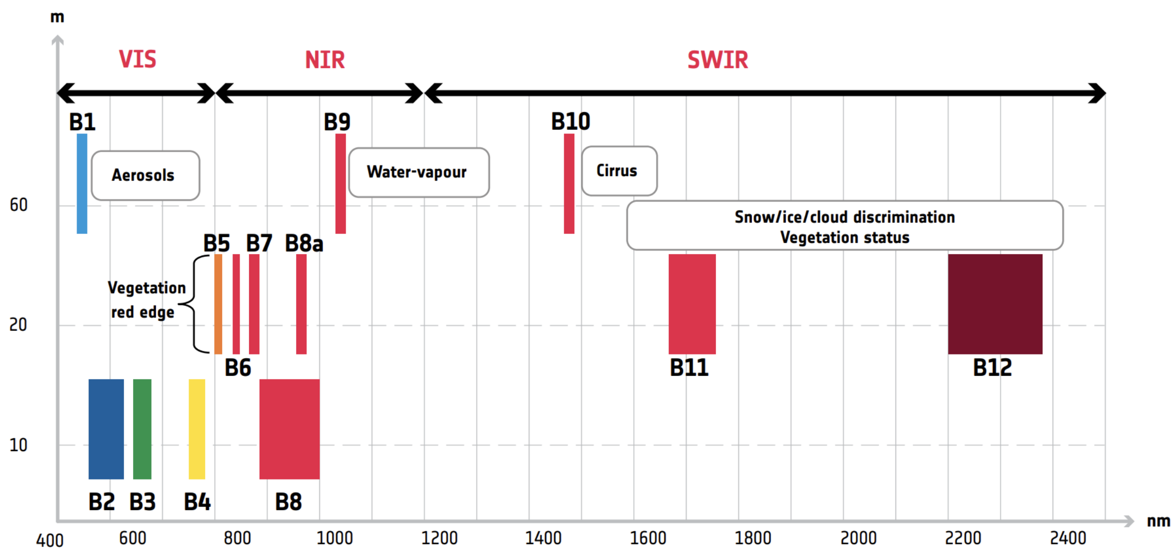

Figure 2.5: Spectral Bands of Sentinel-2. Retrieved from [30]

Sentinel-2 provides high-resolution optical imagery across multiple spectral bands, making it valuable for land monitoring, agriculture, and water quality assessment [119]. The satellite operates in 13 spectral bands with varying resolutions:

- 10 m resolution: Bands 2 (Blue, 490 nm), 3 (Green, 560 nm), 4 (Red, 665 nm), and 8 (NIR, 842 nm) [125];
- 20 m resolution: Bands 5 (Red Edge 1, 705 nm), 6 (Red Edge 2, 740 nm), 7 (Red Edge 3, 783 nm), 8A (Narrow NIR, 865 nm), 11 (SWIR 1, 1610 nm), and 12 (SWIR 2, 2190 nm) [119];
- 60 m resolution: Bands 1 (Coastal Aerosol, 443 nm), 9 (Water Vapour, 945 nm), and 10 (Cirrus, 1375 nm) [71].

2.3.3.2 Landsat

The Landsat programme, managed by NASA and the USGS, has been instrumental in Earth observation since its inception in 1972 [143]. Each mission has contributed uniquely to monitoring the planet's surface, offering invaluable insights into land use, environmental changes, and resource management. In total, nine missions have been carried out to launch the different satellites into orbit [142].

Landsat 1 was named initially Earth Resources Technology Satellite (ERTS). It was the first satellite to provide multispectral imagery of Earth's surface, with a focus on agricultural and environmental monitoring. Launched in 1972 and concluded its mission in 1978 [72].

Landsat 2 continued the mission of Landsat 1, improving the understanding of Earth's resources with its Multispectral Scanner (MSS) and increased reliability. It was sent into orbit from 1975 to 1982 [55].

Landsat 3 introduced thermal imaging capability but faced technical challenges that limited its operational lifespan. It provided critical data on urban growth and agricultural changes. It began operation in 1978 and finished in 1983 [146].

Landsat 4 marked a technological shift with the introduction of the Thematic Mapper (TM), offering improved spatial, spectral, and radiometric resolution for more detailed land surface studies. Starting its functions in 1982 through 1993 [34].

Landsat 5 is a record-breaking mission for operational longevity, providing high-quality data for almost three decades. It enabled extensive monitoring of deforestation, desertification, and water quality. The plan started in 1984 and lasted until 2013 [22].

Landsat 6 designed to carry an Enhanced Thematic Mapper (ETM) for higher resolution imaging, it failed to reach orbit, in 1993, due to a launch vehicle malfunction [142].

Landsat 7, equipped with the Enhanced Thematic Mapper Plus (ETM+), continues to deliver global coverage despite a partial scan line corrector failure in 2003. Data from this mission remains crucial for long-term environmental studies. This critical mission launched in 1999 and remains in Earth's orbit to this day [34].

Landsat 8 (2013–present): Introduced the Operational Land Imager (OLI) and TThermal Infrared Sensor (TIRS), providing improved imaging capabilities and higher data quality, together with the ability to measure surface temperature, providing insights into thermal dynamics and land surface emissivity. It focuses on monitoring land-use changes, vegetation health, and water resources. It got underway in 2013 and remains in orbit around our planet [84].

Landsat 9, being the latest mission, continues the legacy of Landsat 8 and offers high-resolution imagery with advanced radiometric performance to ensure consistency in the long-term dataset. It supports applications in agriculture, forestry, and climate research. The newest innovations in satellite technology that this mission presents, launched in 2021 and is still operational [72].

The bands of the Landsat satellites have different spectral waves [93]. In addition, a figure shows the spectral bands of the Landsat satellites in 2.6, defined below:

Band 1 , also known as the Coastal/Aerosol band, operates at a wavelength of 443 nm and is primarily used for coastal and aerosol studies. It enhances water penetration, making it valuable for bathymetry (water depth estimation) and coastal monitoring. Additionally, it assists in atmospheric correction by reducing scattering effects.

Band 2 , referred to as the Blue band, covers wavelengths between 450 and 515 nm and serves as the blue channel in RGB composites. It helps map water bodies and distinguish soil from vegetation. This band also helps identify man-made structures and analyse water turbidity.

Band 3 , known as the Green band, operates within the 525 to 600 nm range and functions as the green channel in RGB images. It enhances vegetation analysis by capturing strong green reflectance. This band is also beneficial for assessing plant health and monitoring water quality, particularly for detecting oil spills.

Band 4 , the Red band, works within the 630 to 680 nm range and acts as the red channel in RGB images. It is effective for distinguishing vegetation types and monitoring plant stress. Since chlorophyll strongly absorbs red light, this band is particularly useful for vegetation mapping.

Band 5 captures near-NIR light between 845 and 885 nm, making it highly reflective for vegetation. This characteristic makes it essential for biomass estimation and crop health monitoring. It also helps identify water bodies due to the strong absorption of near-NIR light by water, and it is widely used in vegetation indices.

Band 6 belongs to the SWIR region, operating between 1560 and 1660 nm. It is sensitive to plant moisture content and soil composition, helping to differentiate between dry and wet vegetation. This band is also useful for burn scar detection and geological studies.

Band 7 , another SWIR band, captures wavelengths between 2100 and 2300 nm. It is commonly used for detecting moisture in soil and vegetation, making it valuable for lithological and mineral mapping. Additionally, it helps differentiate between snow and clouds.

Band 8 is a panchromatic band covering a broad spectrum of visible light between 500 and 680 nm. It provides high-resolution imagery, enhancing spatial detail when combined with other bands through pan-sharpening techniques.

Band 9 is the cirrus band, operating at wavelengths between 1360 and 1390 nm. It is designed explicitly for cirrus cloud detection, helping improve atmospheric corrections in satellite images by identifying thin cloud cover.

Band 10 and Band 11 capture thermal Infrared (IR) radiation between 10.6 and 12.5 µm. These bands measure surface temperature and heat emissions, making them essential for monitoring land surface temperature, detecting wildfires, and analysing urban heat islands.

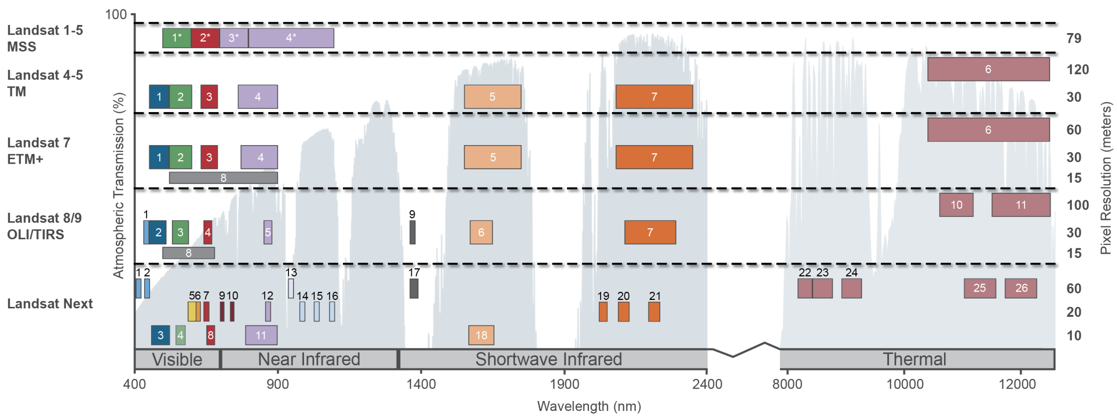

Figure 2.6: Spectral Bands of Landsat. Retrieved from [93]

Landsat has a sixteen-day revisit period, less than the Sentinel-2 but still a valuable and useful beacon of information [84], and has the following resolutions specific to each band:

- 15 meters: Panchromatic Band (Band 8) [72];
- 30 meters: Visible, NIR, and SWIR Bands (Bands 1–7) [93];
- 100 meters: TIRS Bands (Bands 10 and 11) [72].

Table 2.2 resumes all the bands that exist across the most used satellites from the Copernicus Programme and the USGS one. The entries directly match each other where available.

Band	Sentinel-2 (Wavelength & Resolution)	Landsat 8/9 (Wavelength & Resolution)	Description
Coastal Aerosol	443 nm (60 m)	443 nm (30 m)	Used for atmospheric correction and coastal studies.
Blue	490 nm (10 m)	482 nm (30 m)	Penetrates water well, useful for vegetation and soil discrimination.
Green	560 nm (10 m)	561 nm (30 m)	Highlights vegetation, turbid water contrast, and oil spills.
Red	665 nm (10 m)	655 nm (30 m)	Used for vegetation analysis and urban mapping.
Red Edge 1	705 nm (20 m)	–	Sensitive to chlorophyll content in plants.
Red Edge 2	740 nm (20 m)	–	Useful for vegetation health studies.
Red Edge 3	783 nm (20 m)	–	Enhances vegetation classification.
Near Infrared (NIR)	842 nm (10 m)	865 nm (30 m)	Used for biomass monitoring and shoreline mapping.
Narrow NIR	865 nm (20 m)	–	Provides better vegetation differentiation.
Short-wave Infrared (SWIR-1)	1610 nm (20 m)	1608 nm (30 m)	Helps assess moisture content and vegetation stress.
Short-wave Infrared (SWIR-2)	2190 nm (20 m)	2200 nm (30 m)	Differentiates snow and clouds, detects burned areas.
Panchromatic	–	590 nm (15 m)	Higher resolution for sharper images.
Thermal Infrared (TIRS-1)	–	10895 nm (100 m)	Measures land surface temperature.
Thermal Infrared (TIRS-2)	–	12005 nm (100 m)	Useful for detecting heat variations.
Water Vapour	945 nm (60 m)	–	Estimates atmospheric water vapour.

Table 2.2: Comparison of Sentinel-2 and Landsat 8/9 bands, including wavelength and spatial resolution.

2.3.4 Spectral Indices

Spectral indices are mathematical formulas combining reflectance values from specific bands to emphasise certain features or conditions [20]. These indices help researchers extract actionable insights from satellite data. All the following Equations are given with the bands based on Sentinel-2.

Normalized Difference Vegetation Index (NDVI), where its most important use case is to highlight the vegetation in a given image [7]:

$$NDVI = \frac{B_{08} B_{04}}{B_{08} + B_{04}} \tag{2.1}$$

Normalized Difference Water Index (NDWI), mainly used to detect water bodies and evaluate vegetation water content [7].

$$NDWI = \frac{B_{03} - B_{08}}{B_{03} + B_{08}} \tag{2.2}$$

NDCI is useful for estimating Chl-a concentrations in aquatic systems [119].

$$NDCI = \frac{B_{05} - B_{04}}{B_{05} + B_{04}} \tag{2.3}$$

Floating Algae Index (FAI) is effective for identifying harmful algal blooms [96].

$$\text{FAI} = (B_{08} - B_{04}) + \frac{(B_{11} - B_{04}) \times (\lambda B_{08} - \lambda B_{04})}{\lambda B_{11} - \lambda B_{04}} \tag{2.4}$$

Normalized Difference Turbidity Index (NDTI) is applied in assessing water clarity and sediment levels [26].

$$NDTI = \frac{B_{04} B_{03}}{B_{04} + B_{03}} \tag{2.5}$$

Normalized Difference Snow Index (NDSI) is mainly used to accentuate the snow cover over land areas [36].

$$NDSI = \frac{B_{03} - B_{11}}{B_{03} + B_{11}} \tag{2.6}$$

Normalized Burned Ratio Index (NBRI) is a very essential index to estimate the ratio of burned forest through fires [29].

$$NBRI = \frac{B_{08} - B_{12}}{B_{08} + B_{12}} \tag{2.7}$$

These indices are vital for monitoring environmental conditions, supporting applications in agriculture, water quality management, and disaster response. Especially helpful in the short response time of the different phenomena.

2.4 Intelligent Systems

IS is a transformative field of computer science that empowers machines to simulate human intelligence, enabling them to perform tasks such as decision-making, problem-solving, and pattern recognition [61]. In the $21st$ century, IS has emerged as a cornerstone technology, reshaping industries like healthcare, finance, transportation, and environmental monitoring. By automating processes and enhancing efficiency, IS drives innovation, addresses complex challenges, and unlocks unprecedented possibilities in various fields [95].

The evolution of IS has been marked by groundbreaking advancements in algorithms, computational power, and data availability [61]. With its ability to augment human decision-making, IS enhances productivity while fostering novel approaches to tackling societal and technological problems. For instance, IS enables real-time analysis in disaster management, precision medicine, and sustainable urban development, demonstrating its transformative impact [95].

IS refers to the development of systems capable of mimicking cognitive processes such as learning, reasoning, and adapting to new environments [61]. These systems rely on advanced algorithms and computational models to analyse data, uncover patterns, and make informed decisions or predictions. A key strength of IS lies in its ability to process vast datasets and extract actionable insights with precision and reliability, often surpassing human capabilities in terms of consistency, scalability, and speed [95].

Historically, IS has evolved along two main paths: symbolic AI and ML [38]. Symbolic AI, the foundation of early IS research, focuses on explicit rules and logic to perform tasks. In contrast, ML, the dominant paradigm in modern IS, empowers systems to learn from data without requiring explicit instructions. This shift has been driven by advances in DL, a subfield of ML that leverages Artificial Neural Networks (ANNs) to tackle highly complex and data-intensive problems. An illustration, Figure 2.7, of how each field corresponds to the others [67].

2.4.1 Machine Learning

ML is a pivotal subset of IS that enables systems to learn and improve from experience without being explicitly programmed [38]. By analysing large datasets to identify patterns and relationships, ML algorithms perform tasks such as predictions, classifications, and clustering with remarkable accuracy. Unlike

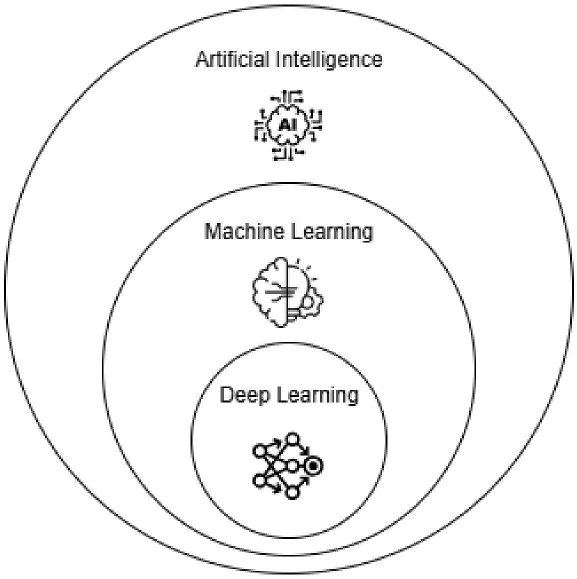

Figure 2.7: Intelligent Systems Fields. Retrieved from [49]

traditional programming, where explicit rules are defined, ML systems adapt to new challenges by deriving insights from examples and their datasets, enhancing their versatility across various applications [67].

The ML process typically involves three key stages: training, validation, and testing [98]. During training, the model learns patterns from historical data. Validation ensures the model's effectiveness on unseen data, while testing evaluates its performance on entirely new datasets. This iterative process refines the model's accuracy, robustness, and generalisation to new scenarios [38].

2.4.2 Learning Paradigms

ML is founded on three core learning paradigms: supervised [98], unsupervised [97], and RL [11], illustrated on Figure 2.8. These paradigms define the approach to model training and problem-solving, each tailored to specific types of data and objectives. Understanding these paradigms is essential for selecting appropriate methods to address diverse challenges and optimise performance in ML applications [38].

2.4.2.1 Supervised Learning

Supervised Learning (SL) involves training a model on labelled datasets, where each input is paired with its corresponding output [98]. A labelled dataset includes examples with input features and their expected outcomes. For instance, in predicting house prices, inputs might include features such as size and location, while the labels indicate the actual prices. The model learns to map inputs to outputs by minimising the error between predictions and true labels. SL is widely used for classifying medical images to detect diseases using labelled datasets, or for predicting loan defaults based on customer profiles in the financial sector [38].

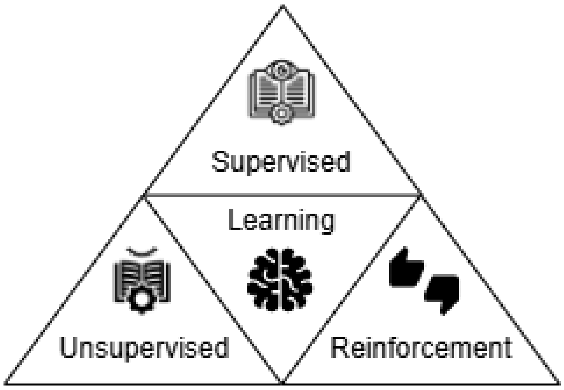

Figure 2.8: Machine Learning Paradigms. Retrieved from [25]

2.4.2.2 Unsupervised Learning

Unsupervised Learning (UL) works with unlabelled data, tasking the model with uncovering hidden patterns or structures [97]. Unlike SL, there are no predefined outputs. Techniques such as clustering group similar data points, while dimensionality reduction simplifies complex datasets. Examples include clustering techniques for segmenting customers in marketing campaigns or detecting fraud patterns in financial transactions [38].

2.4.2.3 Reinforcement Learning

RL involves an agent learning to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties. The agent aims to maximise cumulative rewards over time. RL emphasises sequential decision-making, where each action affects future states and outcomes [11]. A common example is training a robot to navigate a maze, where it earns rewards for progress and penalties for collisions [88]. Another example would be AlphaGo, developed by DeepMind, which learned optimal strategies to defeat human players in the game of Go. It also powers autonomous vehicle navigation and energy management in smart grids [60]. RL has applications in robotics, gaming, and autonomous systems.

2.4.3 Deep Learning

DL is a specialised subset of ML that utilises ANNs with multiple layers, also called deep networks, to model and learn intricate data relationships [67]. This phenomenon is inspired by the structure and functioning of the human brain, in which different neurons are activated to reach a final solution. DL excels at handling structured and particularly unstructured data such as images, audio, and text [38].

A defining feature of DL is its ability to automatically extract features, eliminating the need for manual engineering. By leveraging large datasets and powerful computational resources, Deep Neural Networks (DNNs) achieve state-of-the-art performance in tasks like image recognition, Natural Language Processing

(NLP), and speech synthesis [1, 101]. DL's adaptability and scalability make it a prominent technology of cutting-edge IS.

Recent advancements in DL architectures, such as CNNs [78] and transformers [80], have revolutionised fields ranging from computer vision to language modelling. These models not only process data with very high accuracy but also enable real-time decision-making in areas like autonomous vehicles, predictive maintenance, and personalised healthcare [6, 77, 129].

Over the years, DL has seen many use cases and profoundly helped in tasks that only humans could handle. The following lists showcase some examples:

Image Processing : CNNs detect objects in satellite imagery, such as deforestation or urban growth. They are also integral to facial recognition, medical imaging, and autonomous vehicle systems [129].

Natural Language Processing : Transformer-based models such as Generative Pre-trained (GPT) and Bidirectional Encoder Representations from Transformers (BERT) enable text generation, sentiment analysis, and language translation. Virtual assistants use these models to interpret and respond to user queries, making interactions more intuitive and efficient [1].

Healthcare : DL models assist in early disease detection, such as identifying cancer in medical imaging. They also support drug discovery and personalised medicine by analysing patient data to recommend tailored treatments [77].

DL continues to expand the boundaries of IS, offering innovative solutions to some of humanity's most pressing challenges. Additionally, it can be used to address problems such as water pollution by leveraging the strengths of satellite imagery and computer vision [67].

2.4.4 Transfer Learning

TL is a powerful and widely adopted ML methodology within the DL paradigm. As its name suggests, the core principle is to transfer knowledge gained from solving one problem and apply it to a different but related problem [63]. In practice, this typically involves using a model architecture pre-trained on a large-scale benchmark dataset.

The process involves two primary strategies: feature extraction and fine-tuning. In the feature extraction approach, the convolutional base of the pre-trained model, the series of layers responsible for learning hierarchical features such as edges, textures, and shapes, is preserved. These layers are typically frozen, meaning their weights are locked and not updated during subsequent training [50]. This frozen base acts as a sophisticated feature extractor for the new, smaller dataset. A new classifier layer, typically a few dense layers, is then attached to the base and trained from scratch on the target task.

This technique is exceptionally beneficial for several reasons. First, it significantly reduces the computational resources and time required for training, since the most complex part of the network does not need to be trained from scratch. Second, it democratises the use of DL for tasks where labelled data is scarce [63]. DL models trained from scratch require vast amounts of data to learn meaningful representations, where TL circumvents this by leveraging the robust features already learned from the large source dataset, enabling high performance even with limited data [50].

However, applying TL is not without its challenges. A primary drawback, often called negative transfer, can occur when the source and target domains are too dissimilar. If the features learned by the original model are not relevant to the new task, their presence can hinder rather than improve performance, leading to a poor, unusable model [63]. Furthermore, there is a significant risk of overfitting, especially during fine-tuning. If the new layers are trained too aggressively on a small dataset, the model may become overly specialised to the training data, failing to generalise to new, unseen examples.

Modern methods often involve TL, where pre-trained models are fine-tuned on specific datasets. This approach reduces computational cost and improves accuracy, especially when data is limited. Additionally, ensemble methods combine multiple models to enhance prediction reliability.

2.4.5 Future Directions in Intelligent Systems

As Intelligent Systems continue to advance, two emerging directions stand out as transformative: Multi-Agent System (MAS) and Agentic AI. These approaches expand the capabilities of traditional ML and DL methods by emphasising collaboration, autonomy, and distributed intelligence.

MAS are computational systems composed of multiple interacting agents, each with the ability to perceive, reason, and act within an environment [70]. By distributing tasks among agents, MAS enable complex problem-solving through cooperation, competition, or negotiation. Applications include traffic management, where vehicles act as agents to optimise flow, and environmental monitoring, where distributed sensor agents collaboratively track pollution levels or ecosystem changes [15]. The distributed and decentralised nature of MAS enhances scalability, robustness, and adaptability, making them a key paradigm for next-generation IS [70].

Agentic AI represents a novel evolution of IS, focusing on autonomous decision-making agents capable of reasoning, planning, and executing tasks with minimal human intervention [145]. Unlike conventional AI systems that operate within predefined constraints, Agentic AI emphasises long-horizon autonomy and self-directed problem-solving. These agents can integrate perception, learning, and reasoning to achieve complex objectives across dynamic environments. Recent developments include autonomous scientific discovery, where agents generate hypotheses and design experiments, and adaptive decision-making in domains such as personalised healthcare or smart energy grids [47]. By combining autonomy with adaptability, Agentic AI marks a shift toward systems that not only support but also actively collaborate with humans to achieve strategic goals.

Explainable Artificial Intelligence (XAI) is another critical direction in advancing IS, aiming to make the decision-making processes of complex ML and DL models more transparent, interpretable, and trustworthy [27]. As intelligent systems increasingly influence high-stakes domains such as healthcare, finance, and autonomous driving, the ability to understand why an AI system made a particular decision becomes essential for accountability and user trust. XAI techniques provide human-interpretable explanations without sacrificing predictive power, enabling specialists to validate, contest, or refine system outputs. Emerging methods include attention visualisation in deep networks, causal reasoning frameworks, and post-hoc explanation models like LIME and SHAP. Beyond technical interpretability, XAI supports ethical and regulatory compliance by aligning AI outputs with human values and societal norms. By bridging the gap between black-box intelligence and human understanding, XAI strengthens the usability and acceptance of next-generation IS, ensuring they remain not only powerful but also transparent and responsible [33].

2.5 Image Classification

Image classification is a fundamental task in the field of IS and computer vision. It involves categorising images into predefined classes based on their visual content. This process enables machines to interpret and organise visual data, bridging the gap between raw pixel information and actionable insights. Image classification serves as a cornerstone of numerous applications, ranging from facial recognition and medical diagnosis to satellite image analysis and autonomous navigation [78].

2.5.1 Methods, Techniques & Examples

At the IS level, image classification typically leverages ML and DL algorithms. The process begins with collecting and preprocessing image data to ensure consistency and quality. Preprocessing steps may include resizing, normalisation, and augmentation to enhance model robustness [18].

ML approaches, such as Support Vector Machine (SVM) or Decision Trees (DTs), traditionally required manual feature extraction, in which domain experts identified relevant characteristics such as texture, shape, or colour. However, the advent of DL revolutionised this process by automating feature extraction. CNNs, a type of DL model, excel at image classification tasks due to their hierarchical structure, which captures spatial relationships and complex patterns within images [78].

A significant challenge in image classification is data imbalance, where some classes are underrepresented compared to others. This often leads to biased models that perform well on the majority classes but poorly on the minority ones. Techniques such as data augmentation, re-sampling, and class-weighted loss functions are commonly used to mitigate this issue [4].

Image classification is characterised by several distinct aspects, such as Feature Extraction, which depends on the model's ability to extract meaningful features from images [78]. DL models automatically learn hierarchical features, from simple edges to complex patterns. Another distinct aspect is the scalability of image classification systems, which must handle diverse datasets with varying image sizes, resolutions, and classes. Scalability is crucial for adapting to real-world applications. Another key characteristic is achieving high classification accuracy, which requires balancing the model's complexity and the quality of the training data [98]. Overfitting is a common challenge when models become too specialised in the training data. Finally, interpretability is a concern for DL models, which often act as "black boxes." Efforts to improve interpretability, such as using Grad-CAM to visualise attention maps, are becoming increasingly important. Training and deploying classification models require substantial computational resources, particularly for DL architectures. Optimising efficiency is critical for practical deployment [38].

Diverse examples highlight how image classification is applied across specialised industries [6]. Automated classification of medical images, such as X-rays or Magnetic Resonance Imagings (MRIs), helps detect conditions like pneumonia, cancer, and diabetic retinopathy in healthcare [117]. Satellite and drone imagery classification identifies crop health, pest infestations, and land usage, enabling precision agriculture [73]. Image classification enables autonomous vehicles to recognise traffic signs, pedestrians, and other road elements, contributing to safe navigation [129]. In environmental monitoring, satellite image classification tracks deforestation, urban expansion, and natural disasters, aiding conservation efforts [119]. Image classification also supports product recognition, inventory management, and visual search engines in e-commerce platforms, improving autonomy in retail [68].

2.5.2 Image Classification Models

Image classification models are at the core of many IS, enabling machines to analyse and categorise visual data with remarkable accuracy [38]. These models have revolutionised fields ranging from healthcare to environmental monitoring by leveraging advanced computational architectures to process and interpret images. This section provides an overview of these models, delving into their foundational structures, specific architectures, mathematical formulations, and real-world applications.

Image classification models are predominantly based on CNNs. These are a class of DL models specifically designed to handle image data, exploiting the spatial and hierarchical structure of images to extract features efficiently [78]. The primary advantage of CNNs lies in its ability to automate feature extraction, reducing the need for manual intervention and improving accuracy in complex tasks. Another class of models used for image classification are Transformers, initially developed for NLP but quickly adapted to vision tasks [103].

Among the most influential CNN-based and Transformer models are:

Alexander Network (AlexNet) : A pioneering architecture that demonstrated the power of CNNs by achieving significant performance improvements in the ImageNet competition. It introduced innovations like Rectified Linear Unit (ReLU) activation and dropout [74];

Visual Geometry Group Network (VGGNet) : Known for its simplicity and depth, VGGNet uses small convolutional filters (3×3) and deep architectures to achieve high accuracy in image classification tasks [130];

Residual network (ResNet) : Introduced to address the vanishing gradient problem in deep networks, ResNet uses skip connections to allow gradients to flow directly through layers, enabling the construction of very deep networks [57];

Dense Convolutional Network (DenseNet) : This model connects each layer to every other layer in a feed-forward manner, promoting feature reuse and reducing the number of parameters [62];

Visual Transformer (ViT) : ViT applies the transformer architecture, originally developed for NLP tasks, to image processing by dividing images into fixed-size patches and treating them as sequences. The self-attention mechanism enables the model to capture global dependencies between patches, offering a flexible approach that can scale effectively with large datasets, achieving state-of-the-art performance in image classification tasks [32];

Shifted Windows Transformer (Swin) : The Swin Transformer introduces a hierarchical Vision Transformer architecture that partitions images into non-overlapping windows and processes them locally using self-attention. The innovative shifted-window mechanism ensures cross-window connections, efficiently capturing global and local dependencies. Its hierarchical structure enables multiscale feature learning, making Swin highly effective for dense prediction tasks such as object detection and semantic segmentation. Swin is computationally efficient and scalable, providing state-of-the-art performance across diverse vision benchmarks [83].

These models, among others, have established themselves as benchmarks in the field, each introducing novel techniques to improve performance and efficiency.

2.5.3 CNNs vs. Transformers

CNNs and Transformers represent two dominant paradigms in image classification [78, 80]. While CNNs are efficient at capturing local spatial features through convolutional filters, they often struggle to capture long-range dependencies. Transformers, by contrast, leverage self-attention mechanisms to model global relationships between image patches, thereby capturing both local and global patterns effectively [27].

A key difference lies in data requirements and scaling. CNNs can perform well on relatively smaller datasets due to strong inductive biases such as locality and translation invariance [102]. Transformers, on the other hand, are more data-hungry and require very large-scale datasets (e.g., ImageNet-21k, JFT-300M) to achieve competitive performance. However, once trained on sufficient data, Transformers scale more efficiently and often surpass CNNs in accuracy on large benchmarks. This makes Transformers particularly suitable for applications where massive annotated datasets are available, while CNNs remain practical and effective in data-constrained environments [80].

2.5.4 AlexNet

AlexNet pioneered the use of DL in image classification back in 2012 by winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [74]. It achieved a top-5 error rate of 15.3%, while the second-placed model only achieved a top-5 error rate of 26.2%.

This architecture is built with an 8-layer network comprising 5 convolutional layers with 7×7 filters, followed by 3 Fully Connected Layers (FCL), as shown in Figure 2.9. Inside each convolutional layer, it utilises an activation function (ReLU), which was one of the most essential aspects, helping achieve a low error score by speeding up training time and mitigating the vanishing gradient problem. It enables the network to learn non-linear relationships in the inputs [74]. This model also used pooling layers, which helped reduce the spatial dimensions of feature maps and extract dominant features by summarising regions. In the FCL, a novel approach using dropout layers helped reduce overfitting. This was achieved by randomly nullifying neuron outputs during training. This encourages the network to learn redundant representations for everything and, hence, increases the model's ability to generalise [74].

At the time, the use of Graphical Processing Units (GPUs) to train different models was not common, but AlexNet used this approach, which utilised the parallelism performance that GPUs have, whereas Central Processing Units (CPUs) do not.

AlexNet is the foundation for more modern architectures, inspiring subsequent models such as ResNet and VGGNet by highlighting the importance of depth and computational efficiency.

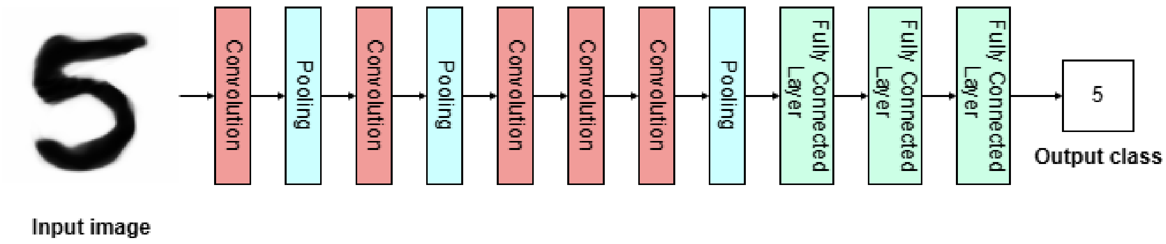

Figure 2.9: AlexNet's Architecture. Adapted from [74]

AlexNet has been widely used for tasks such as defect detection. For instance, AlexNet is a key component of a two-phase method to optimise resource allocation in HetNets. The pre-trained AlexNet architecture is fine-tuned using TL, enabling accurate, computationally efficient scheduling decisions for diverse multimedia services [148].

2.5.5 Inception

In 2014, Inception, also known as GoogLeNet, was published and entered the ILSVRC competition. It won the competition for its efficient use of computational resources while delivering stellar performance. Its selling point was the use of parallelism within a single layer, enabling the model to process features at multiple scales simultaneously [134]. The Inception architecture has undergone many improvements over the years; as a result, many versions are available. Inception version two offers Batch Normalisation, smaller convolutions, more layers, and improved Pooling to enhance training stability and reduce computational costs [64]. The third version increased the model depth with smaller convolutions, but with more layers. It uses regularisation techniques, such as label smoothing, to prevent overconfidence by adjusting the one-hot encoded labels, and uses Root Mean Square Propagations (RMSProps) as the optimiser for better weight updates in training [135]. The fourth and latest version uses modules called Reduction for better grid size reduction [133].

The fourth version starts by feeding the image into the Stem layer, which efficiently downsamples the input and extracts low-level features, as highlighted in Figure 2.10. The downsampled feature maps are then passed to four Inception-A modules to extract features from the input [133]. A Reduction block is then employed, which uses pooling layers and convolutions to reduce the dimensionality of the output from the previous Inception layer. Subsequently, another seven Inception-B modules are implemented to further extract features, followed by a Reduction-B block and an additional three Inception-C layers. The output is fed to an average pooling layer, a dropout layer, and a softmax activation function. At last, for classification purposes, a fully connected layer is used to obtain the output class [3].

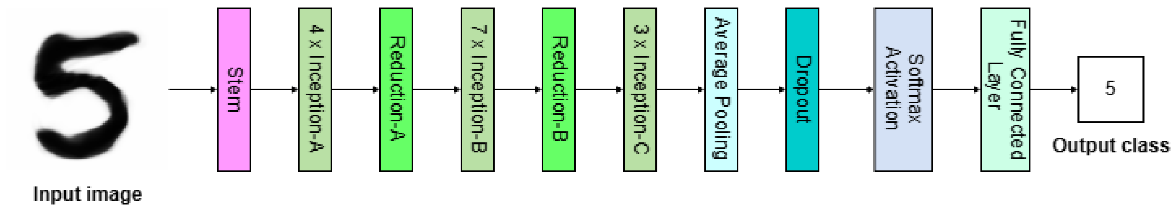

Figure 2.10: Inception's Architecture. Adapted from [133]

According to Figure 2.11, it highlights how the stem layer performs the downsampling of the input image and extracts low-level features from it. It uses a series of 3×3 convolutions and two Max Polling layers, this architecture is employed because it allows to extract low-level features, while computationally being very efficient, instead of using large convolutions. The junction of the parallel paths is then aggregated for subsequent use [133].

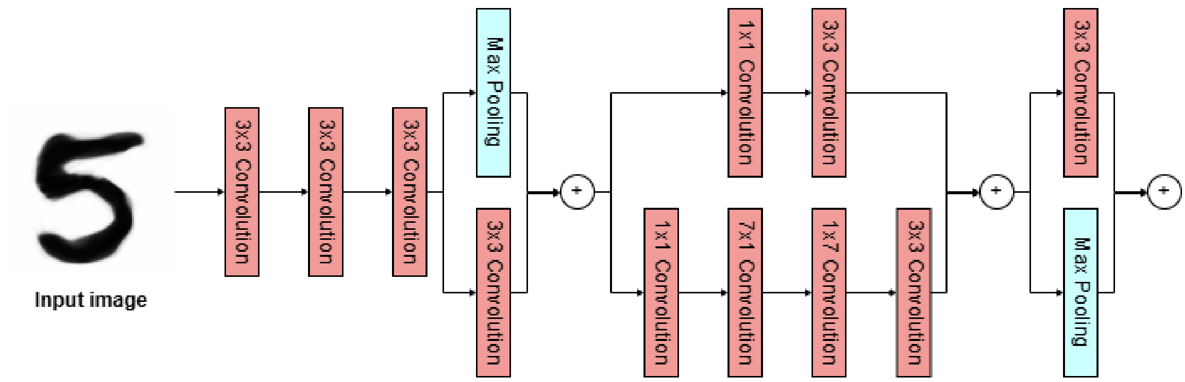

Figure 2.11: Inception's Stem Architecture. Adapted from [133]

The different Inception modules (A, B, and C) differ from each other; for example, Figure 2.12 shows how module A is built [133]. The module follows four different parallel paths to perform feature extraction. 1×1 and 3×3 convolutions are used, with the former downsampled feature maps and the latter extracting features. Average pooling is also used to highlight essential features in a specific region before extracting them. At the end, all is aggregated for further processing. This module has been designed to extract high-resolution image features [3].

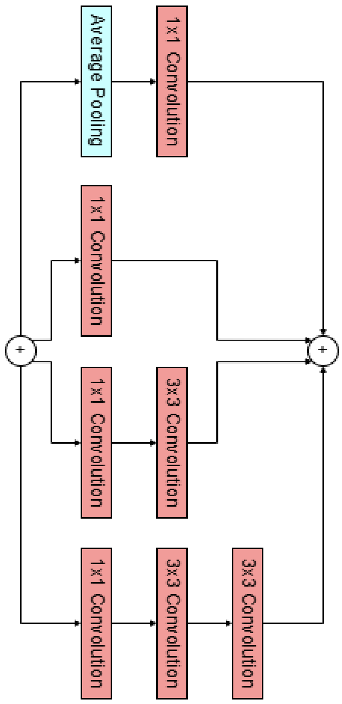

Figure 2.12: Inception's A Module Architecture. Adapted from [133]

The Inception's B module is similar to the previously described module, but with some key differences, as showcased by the Figure 2.13. It again uses 4 parallel paths to perform its work, but it uses different-sized convolutions, such as 7×1 or 1×7. This module is more emphasised for balancing spatial resolution with deeper feature extraction and prioritises efficiency for mid-resolution feature maps, targeting broader spatial scale [133].

The last inception module uses four parallel paths at the start, but when they go deeper, some of them further divide into more paths. When observing Figure 2.14, it finishes with six parallel paths going into the aggregation block. Its purpose serves to extract highly abstract features at lower resolutions, where global patterns dominate. It also focuses on deeper, broader branches to enhance representational power [133].

Inception-v4 is used as a DL model to achieve high accuracy in detecting breast cancer from thermal images, with a modified version, MV4, that improves speed and efficiency by reducing the number of layers

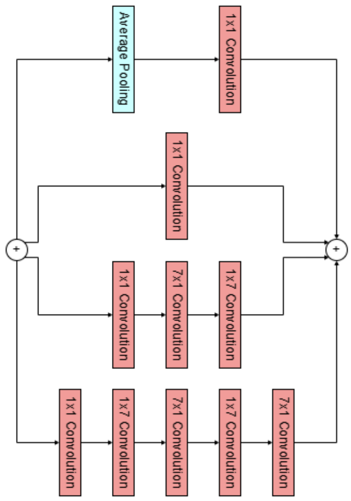

Figure 2.13: Inception's B Module Architecture. Adapted from [133]

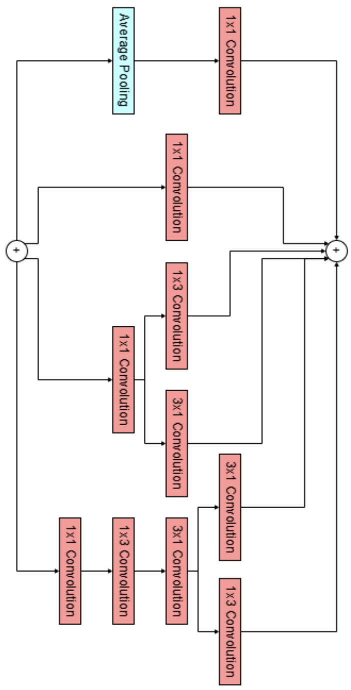

Figure 2.14: Inception's C Module Architecture. Adapted from [133]

[3].

2.5.6 VGGNet

VGGNet was introduced in 2014 and was placed in second place in the ILSVRC competition of the same year [130]. The rationale for its development was to showcase that network depth is critical for achieving better performance in image classification tasks.

Its design is simple and consistent, as exhibited in Figure 2.15, taking much inspiration from the shape of AlexNet's architecture. The difference is that it stacks multiple small 3×3 convolutional filters rather than the 7×7 filter in its counterpart. VGGNet also is much deeper than AlexNet, and the last three FCL' only difference is that each contains more neurons [130].

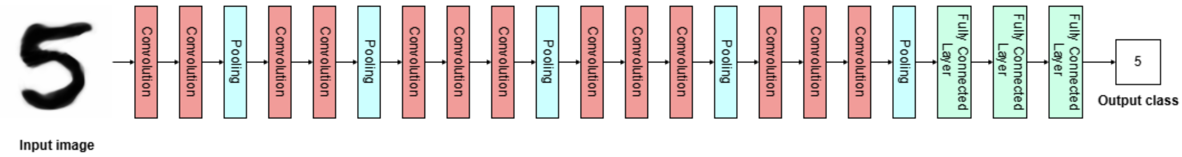

Figure 2.15: VGGNet's Architecture. Adapted from [130]

The advantage of having multiple small convolutional filters instead of larger ones is that it reduces the number of parameters while increasing depth, making the network more efficient at capturing spatial hierarchies and introducing non-linearity. In addition, the added depth allows learning more complex and abstract features, and the added neurons increase the model's complexity, thereby capturing global context effectively [130].

VGGNet's modular architecture makes it ideal for feature extraction. A notable example is its use in agriculture for plant disease classification. It is a powerful tool for plant disease classification using TL, feature extraction, and a layered architecture, achieving high accuracy in distinguishing healthy from diseased plant leaves [106].

2.5.7 ResNet

ResNet won the ILSVRC competition in 2015 with a top-5 error rate of 3.57%. The motivation for designing and implementing this architecture was to address the degradation problem, in which adding more layers to a deep network led to higher training error due to vanishing gradients and difficulties in optimisation [57].

ResNet incorporates basically the same architecture as a conventional CNN, but it introduces the residual connection, so that it can address the vanishing gradient problem [57]. The problem, as mentioned above, occurs during backpropagation when gradients, which measure the change in the error with respect to the parameters, become very small as they propagate through the layers. When gradients are backpropagated to earlier layers, they have little effect on the network because they are so small. Therefore, this architecture implements residual blocks, as shown in Figure 2.16. These blocks implement convolutional layers to extract feature maps, normalisation layers to normalise the inputs for a consistent distribution and activation layers [57].

After the residual block, the gradients are passed to a decision point, where they also receive the gradients of the residual connection, also called the shortcut from before the residual block [57]. It then subtracts the two gradients and passes only the difference to the next block. This allows the network to optimise by focusing on the residual (difference) part of the representation, which becomes increasingly

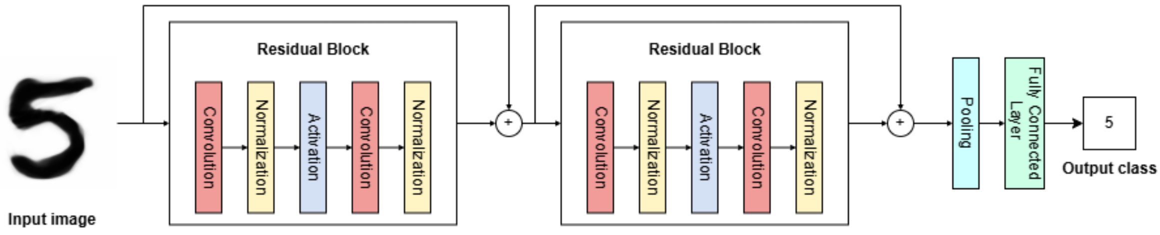

Figure 2.16: ResNet's Architecture. Adapted from [57]

important with really deep networks. These improvements allowed this network to scale up to 152 layers with ResNet-152 [59].

ResNet is considered the next step after AlexNet and has therefore served as inspiration for architectures such as DenseNet and Transformers, particularly in the use of skip connections.

The following Equation explains how the residual block processes the residual incoming information.

$$y = f(x, W) + x \tag{2.8}$$

- $f(x, W)$: Transformation of input x through weights W
- x: Input
- y: Output

The residual block adds the input X to the output of the operations $f(X)$. This shortcut connection allows gradients to flow directly through the identity path, alleviating the vanishing gradient problem. It enables very deep networks to learn effectively [57].

The normalisation layer also plays a vital role in normalising the inputs to achieve a consistent, normal distribution. An Equation has been developed to explain them.

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}}, \quad y = \gamma \hat{x} + \beta \tag{2.9}$$

- μ: Mean of the batch
- σ^2: Variance of the batch
- ϵ: Small constant for numerical stability
- γ, β: Learnable parameters for scaling and shifting
- $\hat{x}$: Normalized input
- y: Final output

Normalisation stabilises the training process by normalising input features to have zero mean μ and unit variance σ^2. Parameters γ and β are learnable, allowing the model to adjust the normalised values. This prevents exploding/vanishing gradients and ensures faster convergence.

ResNet has been adapted for anomaly detection in industrial settings. For example, it has been utilised for quality inspection in manufacturing, providing an accurate and reliable automated system for detecting minor, shallow scratches on plastic casings with complex backgrounds [59].

2.5.8 DenseNet

DenseNet was developed in 2017 and was designed to address inefficiencies in learning and parameter usage seen in deep networks, such as redundancy in feature maps and difficulty in gradient flow [62]. A novel feature, called a dense block, is introduced, as shown in Figure 2.17. Each layer within it connects in a feedforward manner, consisting of a batch normalisation, an activation function, and a convolution layer. This means, for example, that the first layer is connected to all subsequent layers, and the second is then connected to all subsequent layers. Additionally, when subsequent layers receive the feature maps, they will concatenate them rather than sum them. When a dense block has finished processing, to pass the extracted features to the next dense block, a transition layer has been developed, composed of a 1×1 convolutional layer and a 2×2 average pooling layer. It compresses the features from the previous dense block, reducing the feature map size and dimensionality for the next dense block to process. For the purpose of controlling how many feature maps the dense blocks create, a parameter called growth rate is implemented to determine how many feature maps a dense block produces [62].

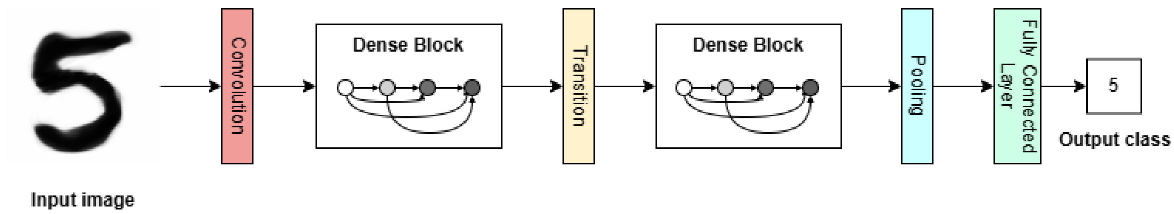

Figure 2.17: DenseNet's Architecture. Adapted from [62]

The benefit of using a dense block is to ensure maximum feature reuse, reduce redundancy, and improve gradient flow. The idea of concatenation, rather than addition, helps preserve low-level information when extracting features from images [62]. The efficiency of DenseNet has enabled networks with shallower layers and fewer parameters than other CNN models, achieving comparable or superior performance. The following formula represents the gradient connections of the dense block.

$$H_\ell = g([H_0, H_1, \ldots, H_{\ell-1}]) \tag{2.10}$$

- H_ℓ: Output of the ℓ-th layer

- $[H_0, H_1, \ldots, H_{\ell-1}]$: Concatenated outputs of all previous layers
- $g(\cdot)$: Transformation (e.g., convolution + activation)

Each layer H_{l+1} concatenates all previous feature maps $[H_0, H_1, \ldots, H_l]$. This ensures maximum information flow between layers and reduces redundancy, leading to efficient feature reuse and stronger gradient flow [62].

DenseNet's efficiency in feature reuse has proven valuable in medical imaging. For example, it has been applied to chest X-ray classification to detect diseases such as pneumonia. It has been used in various medical image classification tasks, including melanoma skin cancer detection, malaria parasite detection, and Alzheimer's disease detection [117].

2.5.9 ResNeXt

Residual Networks with Aggregated Transformations (ResNeXt) was first published in 2017, it is the evolution of ResNet by introducing the concept of grouped convolutions, specifically exploring parallelism and taking advantage from it [144].

Its main building block is the ResNeXt block, where it uses multiple convolutions to perform the feature maps, as depicted in figure 2.18. The process starts by receiving the input image, performing a convolution to extract the first features, and then passing the result to a Max Pooling Layer. The output will then be separated into multiple ResNeXt blocks; depending on the specific model, this could be three or more. In each block, the input is then separated into parallel paths. The number of paths is decided through a parameter called Cardinality [144]. Each patch performs the same operations: starting with a 1x1 convolution for dimensionality reduction, followed by a 3x3 convolution for feature transformation, and, at the end, a 1x1 convolution for dimensionality restoration. All paths emit their outputs to an aggregate point, and the aggregated results are handed off to another aggregate point. In this one, it receives the residual connection coming from the beginning of the block. After the block, another aggregation point collects the output from each block. Depending on the chosen model, there could be a series of blocks after the last described aggregation point to enhance the model's capability to extract features. At the end, the output is passed through a Global Average Pooling layer and then a Fully Connected Layer to perform classification and obtain the output class [144].

The following mathematical expression explains how a ResNeXt block performs the convolutions in parallel:

$$Y = \sum_{g=1}^{C_g} \text{Conv}_g(X_g) \tag{2.11}$$

- C_g: Number of groups (Cardinality);
- X_g: Input to the g^{th} group;

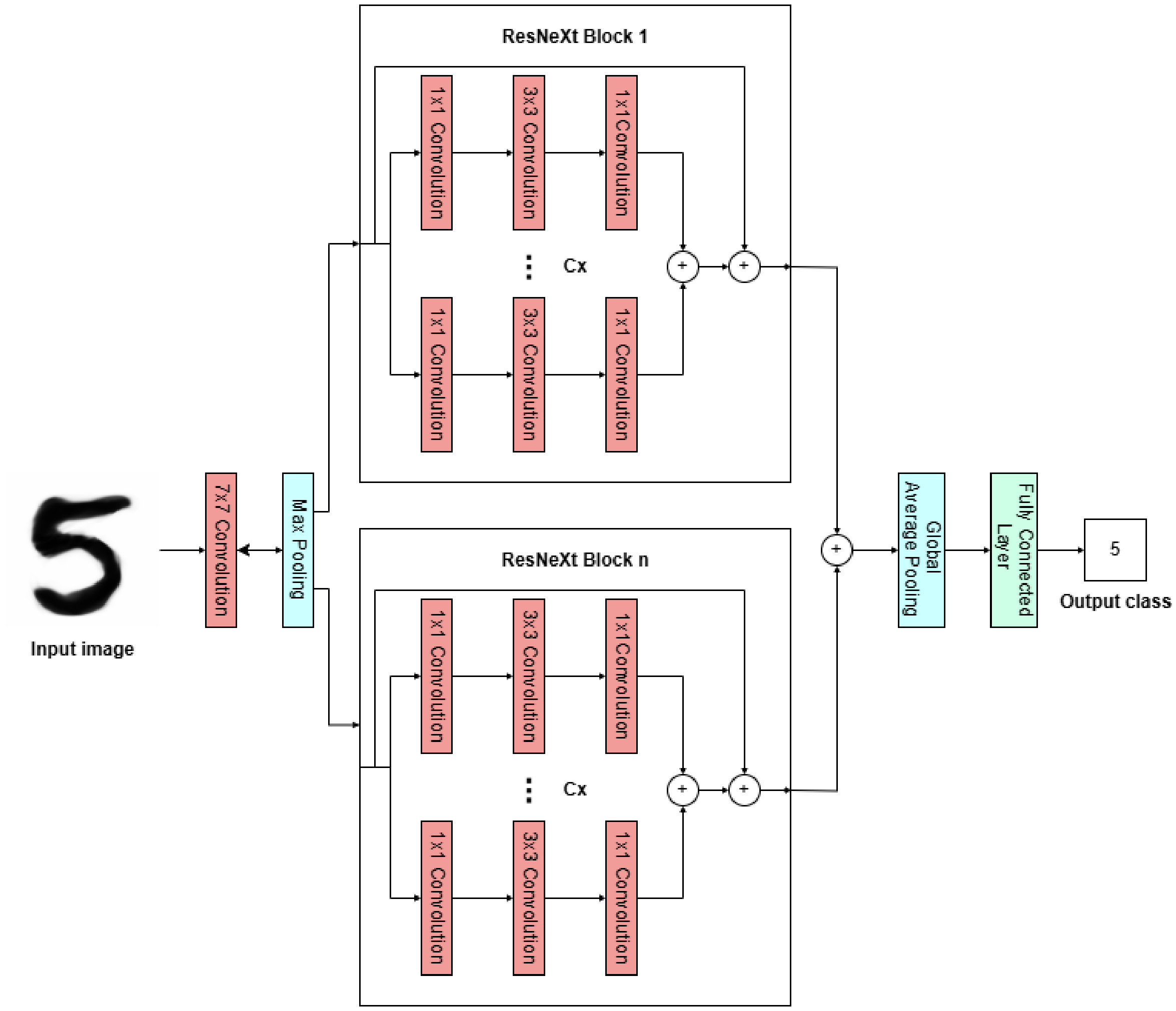

Figure 2.18: ResNeXt's Architecture. Adapted from [144]

- $Conv_g$: Convolution applied to the g^{th} group.

The Equation describes grouped convolution, where the input X is split into C_g groups. Each group g processes its input X_g with a convolution Conv_g, and the outputs are summed to produce Y. This reduces computational costs while maintaining performance, as seen in models like ResNeXt [144].

ResNeXt's benefits include the reduction of the number of parameters and computations compared to traditional convolutional layers, making it more efficient for large-scale image classification tasks [45]. Another benefit is the introduction of cardinality, which allows ResNeXt to capture diverse features across multiple parallel paths, improving the network's representational power. The inclusion of cardinality, depth, and width, which can be adjusted independently, allows it to achieve a balance between computational cost and accuracy [144].

Although ResNeXt reduces computational cost by using parallel paths, memory utilisation will increase. ResNeXt is used for reconstructing transmission images in Tomographic Gamma Scanning (TGS), a nuclear inspection technique widely used to troubleshoot industrial equipment in refineries and petrochemical plants, such as distillation columns and reactors. It is used to reconstruct high-resolution transmission images from sparse projection data obtained from array detectors in TGS [45].

2.5.10 ViT

ViT was launched in 2020, following the success of transformers in NLP, and was adapted for computer vision. It fundamentally differs from a conventional CNN architecture, as it does not use any convolutions to process the input image [32].

Transformers are built to process incoming information as sequences of tokens; therefore, researchers have divided the input image into patches and flattened them into a vector, which occurs in the Linear Projection of Flattened Patches. In contrast, CNNs use convolutional filters; therefore, they can process information at the pixel level, whereas Transformers treat it as a sequence. The sequence has position embeddings together with their corresponding patches, as demonstrated in Figure 2.19.

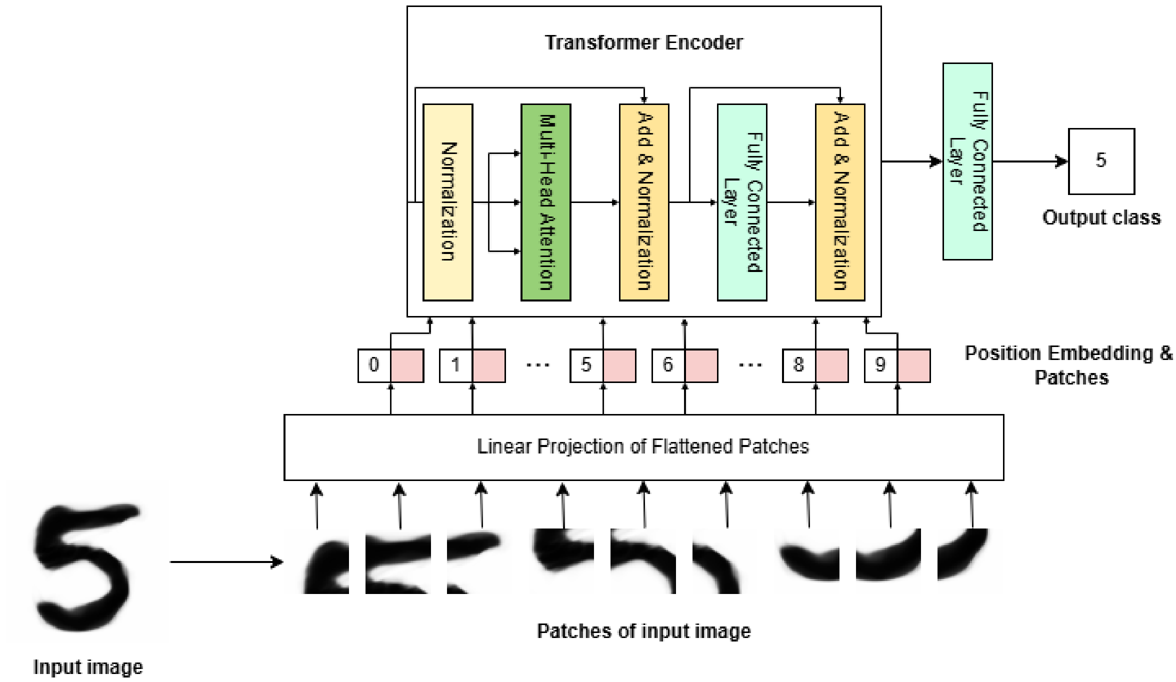

Figure 2.19: ViT's Architecture. Adpated from [32]

It is then passed to the Transformer Encoder, which processes the received embeddings [32]. At the start, it contains a Normalisation layer, which could also be called a pre-normalisation layer; its purpose is to ensure numerical stability and smooth gradients during training. Following up comes the Multi-Head Attention Mechansim (MHA), considered one of the most critical layers in the Transformer architecture. Its application is to compute attention scores across all patches, allowing the model to capture global dependencies between them. Using multiple attention heads enables the model to focus on different aspects of the input. Subsequently, an additional layer is introduced alongside a normalisation function

[32]. The add aspect is inspired by the ResNet architecture, which uses residual connections that connect to the beginning of the residual block. Here, the logic is also applied: the residual connection comes from the start of the encoder block, so the add layer is used to add the initial scores to the processed scores from the MHA. The addition is then normalised to ensure stable training. The upcoming layer is a Fully Connected Layer to provide an additional non-linear transformation of the features. The output is then passed to another Add and Normalisation layer, where the residual connection comes from the input of the previous layer [32]. The Transformer Encoder outputs a special learnable token that is prepended to the sequence of patch embeddings. This token aggregates information from all patches during the self-attention process. These tokens are then passed to a final Fully Connected layer, which predicts the final output class [32].

One advantage of using such a model is that it enables a better global context awareness, because it processes the entire input image as a sequence of patches, allowing the model to capture global dependencies right from the initial layers [32]. It becomes even more pronounced for tasks that require understanding the relationships between distant regions in an image. Another advantage is its lack of bias, which allows it to learn general-purpose representations. A further strong point is that it also works with variable-sized images, as the input is divided into patches, and positional embeddings ensure that spatial information is retained [32].

In summary, this architecture allows researchers to apply other models instead of consistently using CNNs to classify images, enabling them to explore more options based on their requirements. Transformers, having achieved significant success in NLP tasks, can also be explored in computer vision with promising results [73].

Mathematical Equations have been developed to explain the MHA and the Fully Connected Layer to better illustrate the procedure.

Multi-Head Attention Mechanism:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{2.12}$$

- Q: Query matrix
- K: Key matrix
- V: Value matrix
- d_k: Dimension of keys

Last Fully Connected Layer:

$$\text{FullyConnected}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \tag{2.13}$$

- W_1, W_2: Weight matrices
- b_1, b_2: Bias vectors

Output of transformer block:

$$y = \text{FullyConnected}(\text{MultiHead}(x) + x) + \text{MultiHead}(x) \quad (2.14)$$

The Transformer block consists of two parts: MHA and Feedforward Neural Network (FNN). Attention computes relationships between patches by comparing Query (Q), Key (K), and Value (V) matrices. FNN applies non-linearity and transformations to enhance representation learning. $Add\&Norm$ layers stabilize the training process by introducing residual connections [32].

ViT have been used in satellite imagery for land cover classification. A specific study used ViT to segment agricultural fields and urban areas from satellite imagery using the Land Use and Land Cover (LULC) classification. It helped to address critical needs in environmental monitoring, sustainable practices, urban planning, disaster response, forestry, and various other fields [73].

2.5.11 Swin

Swin, presented by Microsoft Research in 2021, addresses challenges of the ViT model, such as training time, while maintaining or surpassing performance [83]. This model was developed to address the limitations of ViT, particularly their inefficiency in handling high-resolution images and their lack of hierarchical feature representation. The Swin Transformer builds on ViT's success while incorporating key innovations, such as shifted windows for efficient computation and stronger cross-window connections. It is a hierarchical architecture that allows multiscale feature extraction. Unlike ViT, which treats the entire image as a single sequence, Swin divides the image into patches organised into non-overlapping windows, enabling a more localised computation approach [83].

When analysing Figure 2.20, one can observe that there are four stages, each with two Swin Transformer Blocks and a Linear Embedding Layer or a Patch Merging Layer. More noticeably, the third stage has six Swin Transformer Blocks. When starting the operation, the input image gets subdivided into small patches, called Patch Partition. It is then handed over to the first stage for processing, in particular to Linear Embedding, which projects the sequence of patches into an arbitrary dimension and converts them into tokens. It is then passed to the first Swin Transformer Block, where it applies a normalisation layer, and the Window-based Multi-Head Self-Attention mechanism then operates on the input [83].

This layer takes a fixed-sized window and applies the attention mechanism only to that window of the input. This locality within each window can be compared to the locality in convolutions, but with the added flexibility of self-attention. The output is then handed over to an Add & Normalisation Layer, where the idea of residual connections has been inspired by the ResNet architecture, where the residual connection comes from the start of the block [57]. Afterwards, a Fully Connected Layer models non-linear transformations. At that point, the first Swin Transformer Block has finished, and it is then transferred to the next Swin Transformer Block. This one implements the same idea as the first block, but the main difference lies in the second layer, which uses the Shifted-Window Multi-Head Self-Attention Mechanism. It models the

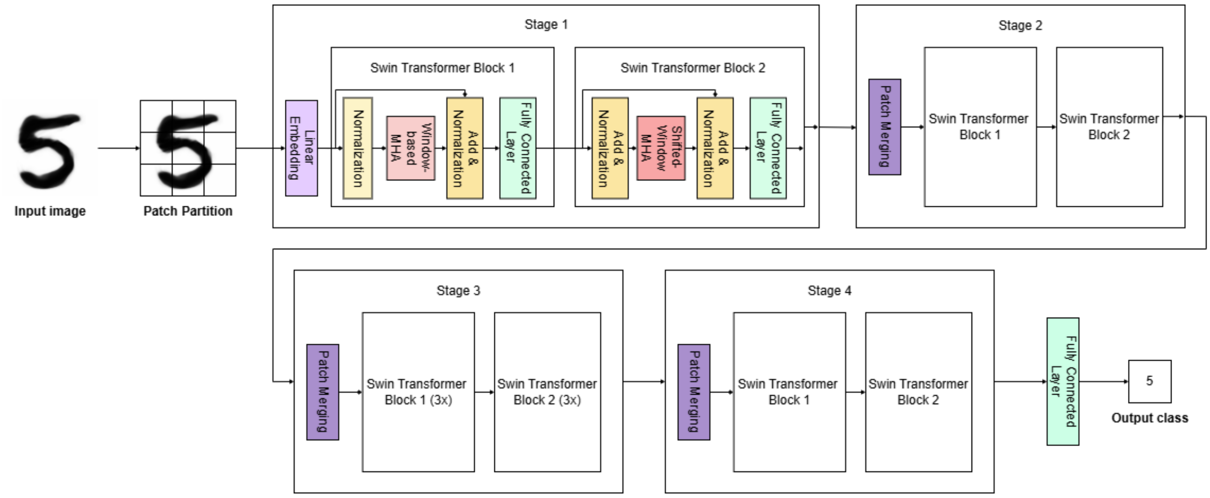

Figure 2.20: Swin's Architecture. Adapted from [83]

relation between windows by shifting the window. The shifting operation usually shifts by half the window size, creating overlaps between each window, capturing global dependencies [83].

Window-Based MHA:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^\top}{\sqrt{d_k}} + B\right)V \tag{2.15}$$

- $Q = XW_Q$: Query matrix;
- $K = XW_K$: Key matrix;
- $V = XW_V$: Value matrix;
- W_Q, W_K, W_V: Learnable Weight Matrices;
- d_k: Dimensionality of Keys;
- B: Relative positional bias matrix encoding spatial relationships within the window;
- Softmax: Normalised attention scores.

Equation 2.15 defines the attention mechanism applied within a local window. The input Q, K, and V matrices, derived using learnable weights W_Q, W_K, and W_V, calculate scaled dot-product attention. A relative positional bias B is added to encode spatial relationships, and the softmax operation normalises attention scores. This mechanism focuses computational resources on smaller, localised regions of the input [83].

Shifted-Window MHA:

$$\text{ShiftedAttention}(X) = \text{ReverseShift}\left(\text{WindowAttention}(S(X))\right) \tag{2.16}$$

- $S(X)$: Cyclic shift operation applied to the input X;
- $WindowAttention$: MHA applied within each window;
- $ReverseShift$: reverses the cyclic shift $S(X)$.

This Equation 2.16 describes the shifted-window multi-head attention. The input X undergoes a cyclic shift $S(X)$ to offset windows, allowing cross-window interactions. Window attention is applied within each shifted window, and the reverse shift restores the original alignment. This improves feature interaction between neighbouring windows [83].

Linear Embedding:

$$Z = XW + b \tag{2.17}$$

- $X \in \mathbb{R}N(PPC)$: Flattened patches, with N as the number of patches, P as the patch size, and C as the number of channels;
- $W \in \mathbb{R}^{(PPC)D}$: Linear projection weight matrix;
- $b \in \mathbb{R}^{D}$: Bias vector;
- D: Dimensionality of the projected embedding space.

Linear embedding's Equation 2.17 projects the flattened input patches X into a higher-dimensional space. The learnable weight matrix W and bias b perform a linear transformation, producing Z, which serves as the model's embedding space for patch representations. This step bridges raw input patches to a meaningful latent space for further processing [83].

In subsequent stages, they perform Patch Merging at the start of it. It reduces the spatial resolution of feature maps while simultaneously increasing the number of feature channels, similar to the Pooling Operation in CNNs, highlighted in Figure 2.21 [83]. The process of merging implies that neighbouring patches are typically grouped together and represented as a single token, as shown in step one. In step two, these grouped patches are then concatenated along the channel dimension, also called stacks, increasing the feature dimension. At the end, the stacks are combined, as shown in step three [83].

Patch Merging:

$$X' = \text{Reshape}(X) \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times (4C)} \tag{2.18}$$

$$Z = X'W + b \tag{2.19}$$

- $X \in \mathbb{R}^{HWC}$: Input feature map;

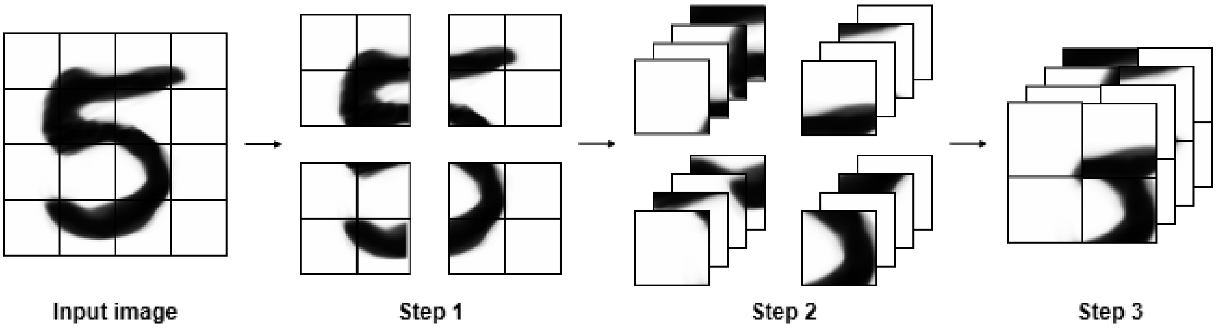

Figure 2.21: Swin's Patch Merging Architecture. Adapted from [83]

- $W \in \mathbb{R}^{(PPC)D}$: Reshaped feature map with $\frac{H}{2} \times \frac{H}{2}$ spatial resolution and $4C$ channels;
- $W \in \mathbb{R}^{(4C)C}$: Linear projection weight matrix for patch merging;
- C': New channel dimensionality after patch merging;
- $b \in \mathbb{R}^{D}$: Bias vector;

Equation 2.18 reshapes the input X into a new format, halving the spatial resolution while concatenating features into 4C channels. In addition, Equation 2.19, the reshaped feature map X' undergoes a linear transformation using the weight matrix W and bias b, resulting in the output Z. This operation enables the model to focus on larger-scale patterns by aggregating patch features.

The benefits start with a much more efficient use of self-attention, enabling effective processing of high-resolution images [83]. Another is to capture local and global dependencies instead of only global ones. The ability to process input features at multiple scales, starting with small patches and progressively merging them into larger ones, creating a hierarchical representation similar to that of traditional CNNs [83].

The Swin Transformer architecture is an improvement over the previous ViT architecture by leveraging enhancements across the entire process and reducing overall computational costs. The Swin Transformer is applied in the Swin transformer based multi-feature integration network (Swin-MFINet) model for pixel-level surface defect detection, enabling it to capture global semantic features while remaining computationally efficient through its shifted-window approach [150].

2.6 Literature Review

This section provides a summary of five studies, selected from a larger set, related to this dissertation. The aim is to understand what has already been done, then to perform a critical analysis of each study to determine whether there is room for improvement. This will then be taken into account for our research.

In the article [120], the authors use different techniques to monitor water quality in inland water bodies, such as lakes and rivers. The locations were in the Midwest United States, specifically Carlyle Lake, located in southwestern Illinois, with an area of 105 km^2. Another location was Lake Decatur, located in central Illinois, with a size of 11 km^2. Finally, the last location was the Meramec River in Missouri. The lakes are mostly used for agricultural purposes, resulting in notable agricultural runoff, increased phosphorus and NO_3 concentrations, and occasional algal blooms. The river was mostly clean because the water was always flowing and human activities were mostly unimpacted, setting the baseline for water quality comparisons. The researchers then selected different water quality parameters to monitor optically active parameters, such as Chl-a, turbidity, Fluorescent Dissolved Organic Matter (fDOM), Blue-Green Algae Phycocyanin (BGA-PC), and Total Suspended Solids (TSS), and non-optically active parameters, such as nitrogen, phosphorus, and Dissolved Oxygen (DO). The latter can only be indirectly estimated because they are not optically active, meaning they cannot be seen in images. Using RS, Landsat-8 and Sentinel-2 satellite images, and in-situ data from proximal sensors, they have built a dataset. The dataset from the proximal sensors came from other studies, particularly from the National Great Rivers Research and Education Centre, from two buoys —floating devices designed to house sensors to collect data, and from data collected by the researchers. The authors combined the two RS datasets to improve the temporal resolution. By filtering for cloud cover and matching with ground sampling locations, they ended up with 96 images. They have used DL models, specifically CNNs and Long Short-Term Memory (LSTM), to combine the spectral and temporal features. They also used Multiple Linear Regression (MLR), Partial Least Squares (PLS) and Support Vector Regression (SVR). Their metrics were Coefficient of Determination (R^2), Median Absolute Deviation (MAD), Mean Square Error (MSE), Root-Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). It outperformed traditional regression models, particularly for optically active parameters, due to its ability to capture non-linear relationships. DL methods achieved overall superior performance, with higher R^2 score and the lowest errors across all metrics. The DNN model showcased the best performance in R^2 with BGA-PC getting 0.82, Chl-a yielding 0.85, TSS attaining 0.84, turbidity acquiring 0.65, NO3-N getting 0.29, PO4-P achieving 0 and TDS achieving 0.36. The RMSE of the same model was in the same order as above: 0.43, 2.31, 3.73, 26.36, 0.31, 0.25 and 48.68. The noted limitations of this study were that the transferability of the models is very complex due to the differences in optical properties and parameter relationships, decreasing the generalisability of the model. Various limiting factors in satellite imagery, such as cloud cover and shadows, reduced the image quality. The non-optical parameter posed a significant challenge due to inconsistent data across datasets. Nonetheless, the researchers provided promising directions in utilising RS in these types of studies, by increasing the comprehensive monitoring between all the different data and an anomaly detection system was proposed for early detection of water quality issues such as harmful algal blooms.

The study conducted by Fangling and Ding et al. [108] explores the use of a CNN to classify water quality in inland lakes using in correlation with RS data. It focuses on developing a cost-effective method for water-quality monitoring using satellite imagery. The location of the study is situated at Erhai Lake and Chaohu Lake, both located in China. The authors used Landsat-8 satellite imagery between 2014 and 2018, summing up to 81 images, eliminating the ones with a high percentage of cloud cover. In-situ data has been collected through official national observation repositories and monitoring sites operated by local governments. The data for Chaohu lake included PH, DO, Chemical Oxygen Demand (COD), Ammonia Nitrogen (NH_3-N) and WQI classification (Class 1, 2, 3, 4 or 5). For the Erhai Lake, it also included Total Phosphorus (TP), Total Nitrogen (TN) and DO. The different WQI classes represent different qualities; the first and second classes represent good water quality grouped in class A, the third class indicates water suitable for fisheries and industry, renamed to class B, the fourth and above classes are not ideal for drinking, but can be used for recreation and irrigation purposes. These classes were grouped to represent class C. The satellite images have been corrected for atmospheric effects to achieve the best possible image quality. The DL model used was a CNN built from the ground up but inspired by the AlexNet architecture. Four convolutional layers, ReLU activation functions, and three FCL were integrated, but pooling layers were removed to preserve information. At the end, a softmax layer was applied to classify the output into one of the three classes. The CNN was trained using a 4:1 ratio of training to test samples, with training epochs set to 1000, a learning rate of 10–5 and momentum of 0.9. The dropout for the first two FCL was set to 50% and overall accuracy was used as the performance metric. The accuracy of the CNN model was 89.90% with two convolutional layers, 92.26% with three convolutional layers, 92.49% with four convolutional layers, 92.40% with five convolutional layers, 92.10% with six convolutional layers, 92.09% with seven convolutional layers. The one with four convolutional layers performed best, based on both accuracy and computational complexity. Limitations of the study included the limiting data availability, therefore TL was used to mitigate this limitation.

The research published in [91] investigates using a CNN based on the AlexNet architecture to classify water quality levels in inland lakes using Sentinel-2 satellite images. The study also examines the relationships among water quality, meteorological factors, and algal blooms. The geographical site explored is in Lake Dianchi, China, which is one of the three most eutrophic lakes in China. For the data acquisition, they used Sentinel-2 images collected from November 2020 to April 2023, in total, 59 images. Atmospheric corrections were applied, and only images with no cloud cover were selected. Water quality data, including TP, TN, WT, and PH, were obtained from an automatic water quality station. It was classified into three levels: Level A (good), Level B (mildly polluted), and Level C (severely polluted). Researchers also included meteorological weather conditions such as Relative Humidity (RH), Precipitation (PP), Atmospheric Temperature (AT), and Wind Speed (WS). A statistical analysis was done to understand which water parameter influenced more the area of algal bloom. The main one was TP, while meteorological factors such as WS, RH, and PP played secondary roles. The DL model developed was based on the AlexNet architecture, due to its simpler design and lower computational complexity. An algal bloom extraction algorithm was developed using the NDVI and FAI. Algal blooms were identified when NDVI > -0.1 or FAI > 0.00. The CNN model is composed of five convolutional layers, where the first one includes a ReLU activation function, one max pooling layer and three FCL. The samples were divided into training and test sets in a 4:1 ratio. The model was trained over 500 epochs with an initial learning rate of 10–4. The performance of the CNN model was evaluated using metrics such as accuracy, precision, recall, and Harmonic Mean of the Precision and Recall (F1-score). Between November 2020 and April 2023, the water quality of Lake Dianchi was classified as Level A (good water quality) in 1.24% of cases, Level B (mildly polluted) in 84.28%

of cases, and Level C (severely polluted) in 14.48% of cases. Hence, despite fluctuations in TP, a recent stabilisation at 0.05 mg/L suggests a favourable course for the future management of algal blooms.

In article [137], the researchers explore a satellite image-based monitoring system using a CNN to assess water quality. The main goal of the system is to provide accurate information on water quality indicators such as PH, temperature, turbidity, salinity, and Chl-a. The location of the study area is the Kavali Lake, situated in India. Satellite images have been gathered from different sources, as Sentinel-2 or Landsat-8. Corrections to these images were applied, such as radiometric and geometric atmospheric ones; normalisation was also applied. The dataset was divided into training, testing and validation. The use of labels may be necessary depending on the interest area. The labelling classifies the water location as either polluted or clean. The development of a suitable CNN involves including convolutional, pooling, fully connected, and output layers, and adjusting the model architecture based on the specific targets. Different WQI parameters were used and their definition of amount was the following, PH between 6.6 and 8.5; salinity more than 1, turbidity between -0.2 and -0.1, Chl-a between -0.1 and 0.1 and temperature between 15 and 35. The accuracy metric was used to assess the model's performance, yielding 98%. A time-series analysis of water quality in Kavali Cheruvu from January 5, 2023, to January 28, 2023, reveals fluctuations in water quality during this period. Water quality improved by January 20, followed by a slight decline by January 28. It concluded that this guideline of constructing the framework helps to quickly and accurately assess water quality over large areas and, over time, can contribute to better management of water resources and more timely interventions to maintain the health of aquatic ecosystems.

The last article, in particular [5], focuses on a framework for water quality classification using a combined CNN and LSTM model that aims to improve water quality prediction. The classification classes boiled down to safe or unsafe. The study implements a CNN as the front-end that processes the input features from the provided images, capturing spatial features. Then, the LSTM acts as a back-end, receiving the CNN 's output to predict whether the water is safe or unsafe. The data has been acquired on a website that hosts a large number of datasets, called Kaggle. The dataset included 21 parameters related to water quality and a final attribute indicating whether it is safe or not. The data has been pre-processed, resulting in values ranging from 0 to 1. A dynamic (Particle Swarm Optimization (PSO)) based on exponential decay and particle elimination has been employed to select the ten best features. The model's architecture is built with convolutional, pooling, and FCL. The LSTM consists of a cell, an input gate, an output gate, and a forget gate to capture temporal dynamics. When combined, the first processes the non-linear characteristics of the input features, and the latter receives the abstracted data to perform classification. The evaluation metrics used to assess the model's performance were accuracy, precision, recall, and F1-score. The authors found that the model's accuracy was 99.99%, exceeding that of both existing classifiers, with and without feature selection. The proposed model outperforms an existing ensemble method, which achieved a classification accuracy of 98.1%. Lastly, the paper concludes that the developed model is highly effective for water quality prediction, achieving notable improvements in accuracy over existing methods by combining dynamic PSO for feature selection with a CNN-LSTM model for classification.

2.7 Critical Analysis

Several studies have proposed methods to classify water quality parameters using DL models. While these approaches demonstrate significant advancements in the field, they also exhibit certain limitations that require careful evaluation. This subsection critically analyses the strengths and weaknesses of key contributions in this area.

Novel findings have been reported in the different articles by applying simple CNN methods, as in [91, 108, 137], or hybrid approaches, as in [5, 120]. The analysis of the performance of all the models has been done using different metrics like [5, 91, 120] used R^2, MAD, MSE, RMSE and MAPE for regression tasks in [120]. In contrast, accuracy, precision, recall, and F1-score were used for classification tasks in [5, 91]. The remaining articles [108, 137] used accuracy alone to benchmark their models, making their lack of understanding hard to grasp in more depth.

In more detail, article [120] highlighted the difficulty of estimating non-optically active parameters, but no proposal was made to improve the model to assess them; hybrid approaches or indirect estimation through mathematical formulas could have been explored. Another limiting factor was the small dataset used, which affected the model's generalisability and led to overfitting; this was not explicitly discussed or addressed. No evaluation metrics or graphs were presented to show whether overfitting was occurring. The addition of other datasets, such as meteorological data, could have been included to determine whether weather affects water pollution.

The following article [108] approached the problem of water pollution by focusing more on the overall water quality instead of focusing on individual parameters affecting it. It used TL to mitigate the small dataset, including preprocessing steps such as atmospheric correction and matching satellite data with in-situ data. However, it does not mention data normalisation. Optimal hyperparameter search via grid search or other techniques was not explored; thus, potential improvements could have been achieved. Although the spatial resolution of Landsat 8 images, which are 30m, is sufficient for large lakes, it might not fully capture local variations in water quality, especially near lake inlets, where most in-situ data were collected. Using higher-resolution imagery might yield more accurate water quality maps for the entire water body, such as Sentinel-2 at 10m.

The third article [91] addressed the issue by utilising a CNN to predict the different parameters of the water to understand algal blooms. It included the utilisation of other datasets than those in situ; it used meteorological datasets such as RH, PP, AT and WS. Unfortunately, the model explored and developed was a simple CNN without any hyperparameter optimisation, thereby also neglecting overfitting analysis. Other techniques to generalise the model, such as data augmentation, were not explored to increment the already small dataset, or the use of cross-validation would also be beneficial. The limit of 21 pixels around the monitoring stations could have been increased to determine whether more distant waters could be polluted.

The fourth article [137] focused on using RS together with CNN to analyse water quality, providing guidelines for such a development task. The guidelines are good, but the analysis of the developed model was very lacklustre, providing only accuracy, and the construction of their framework did not include specific implementation details, such as the number of layers used. It included thresholds to determine whether the water is polluted. The lack of specific metrics to assess the model was also a key component for understanding its performance; it only stated that it had an accuracy of 98%. This performance does not include any overfitting analysis or more in-depth analysis of the data, only mentioning what different parameters were predicted.

The fifth and final article [5] develops a hybrid model based on CNN and LSTM to predict the water quality. Pre-processing was applied: a standard scaler was used to scale the data to the same range, and then a dynamic PSO was used to select the ten best features. Unfortunately, data augmentation techniques were not applied to the data and not providing the number of entries in the dataset. When developing the models, the study does not include any systematic hyperparameter tuning, leaving potential performance on the table. Not only was the dataset size not mentioned, but the inclusion of other datasets could have complemented it, providing the model with more understanding. Data augmentation was also not included in the study. An analysis of overfitting was not included; therefore, it is unclear whether the model was overfitting, especially given the high accuracy of 99.99%.

Articles	Pre-processing of Data	Hyperparameter Tuning	Other Datasets Used (e.g., Meteo)	Overfitting Analysis	Cross-Validation	Data Augmentation
[120]	✓	×	×	×	×	×
[108]	✓	×	×	×	×	×
[91]	✓	×	✓	×	×	×
[137]	×	×	×	×	×	×
[5]	✓	×	×	×	×	×

Table 2.3: Analysis of Articles Based on Various Criteria

This dissertation will address the aforementioned issues by developing a more robust and comprehensive DL-based framework for water quality assessment. Specifically, overfitting will be carefully analysed using validation curves, learning curves, and regularisation strategies, ensuring that the models generalise well beyond the training data. Data augmentation techniques will be applied to expand and analyse the dataset's variability, thereby reducing the risk of overfitting and improving model stability. Furthermore, the integration of additional datasets, including meteorological and hydrological information, will be explored to capture external drivers of water quality fluctuations and to improve classification performance. The dissertation will also incorporate systematic hyperparameter optimisation and cross-validation strategies to ensure fair benchmarking of the models. By addressing these limitations in a structured manner, the framework aims not only to enhance the accuracy and reliability of water quality predictions but also to provide a transparent, reproducible, and scalable methodology adaptable to different water bodies and monitoring contexts.

2.8 Summary

Chapter 2 provided an extensive review of the theoretical and technological foundations relevant to this study. It began by defining the research methodology used to identify and evaluate the most significant works across the domains of environmental sustainability, pollution, IS, and RS. The chapter explored how water pollution affects ecosystems and human health, describing different pollutant categories and their impacts. It then introduced the role of RS in environmental monitoring, detailing major satellite programmes such as Copernicus and USGS, and highlighting the importance of spectral indices in assessing water quality from space-based observations.

The chapter further delved into the core concepts of AI and IS, outlining the progression from traditional ML techniques to DL and RL. It reviewed key DL architectures, including CNNs and Transformers, such as AlexNet, ResNet, DenseNet, ViT, and Swin, evaluating their relevance to environmental and image classification tasks. A critical analysis of prior research identified significant limitations, such as the lack of systematic hyperparameter tuning, insufficient dataset diversity, and limited use of data augmentation. These gaps justified the need for this study's comprehensive and reproducible DL-based framework for water quality classification.

Chapter 3

Materials & Methods

In this chapter, the focus will be on the different aspects of the Materials and Methods. In section 3.1, the study areas will be defined, along with the acquisition of the data. The subsequent section 3.2 will present visualisations of the study area using satellite imagery and of the data, including details. Finally, section 3.3 provides details on the data pre-processing for later use.

3.1 Data Collection

The study locations are two lakes on the *São Miguel* island of the Azores archipelago. The first one being the lake *Azul* and the second one being the lake *Fogo*. These lakes were selected for their ecological importance and for ongoing water-quality monitoring by local authorities.

3.1.1 Lake Characteristics

Lake Azul is situated in the municipality of *Ponta Delgada* within the parish of *Sete Cidades*, and is located at an altitude of 260 metres above sea level [46]. The lake spans a maximum length of 2,670 metres and a width of 2,080 metres. Its surface area measures approximately 3.5 square kilometres, and it is supported by a drainage basin of 15.25 square kilometres. The lake holds a total water volume of 47,361,401 cubic metres and reaches a maximum depth of 28.5 metres. The designated sampling point for data collection is located at coordinates 607721 Easting and 4192612 Northing, according to the World Geodetic System (WGS) 1984 UTM Zone 26N reference system [46].

In contrast, Lake Fogo is positioned at a significantly higher elevation of 575 metres and lies within the municipalities of *Ribeira Grande* and *Vila Franca do Campo*, overlapping the parishes of *Conceição* and *Água D'Alto* [46]. This lake has a maximum length of 2,312 metres and a width of 1,078 metres. With a surface area of 1.59 square kilometres and a basin area of 5.03 square kilometres, Lake Fogo contains a stored water volume of 23,443,191 cubic metres and reaches a maximum depth of 30 metres. Sampling for this lake was carried out at coordinates 634431 Easting and 4180852 Northing, also in the WGS 1984 UTM Zone 26N projection [46].

3.1.2 Tabular Data Acquisition

The local authorities collect samples every year, for each month, from all the lakes in the Azores archipelago. These collections occur every year, and their findings are published every three years [46]. The dataset covers the years 2017-2020. These samples are then sent to a laboratory for testing to assess the quality of the lakes' water. They perform analysis on different parameters such as temperature, DO, PH, conductivity, orthophosphate, TP, NO_3, TN, Biochemical Oxygen Demand (BOD), COD and Chl-a [46].

After the analysis is complete, they classify the water's quality according to the local authorities' rules. They divide the classification into three main categories: ecological, physical-chemical, and hydromorphological rules. The first classification approach is based on the Ecological Quality Ratios (EQR), which is determined by comparing the Chl-a concentration to a reference condition. This is done by first assigning a score to the observed Chl-a value based on predefined ecological thresholds. Specifically, if the Chl-a concentration is less than 3 μg/L, the sample receives five points. If the value is greater than or equal to 3 μg/L but less than 10 μg/L, it receives three points. Finally, if the concentration is greater than or equal to 10 μg/L, it is assigned one point [46].

Once the score is assigned, the score ratio is calculated by dividing the score by a reference value of 4.7, corresponding to the maximum possible score under ideal conditions [46]. This ratio then allows for the ecological status to be categorized into one of five classes: A score ratio greater than 0.94 is classified as Excellent, one between 0.74 and 0.94 is considered Good, values between 0.53 and 0.74 fall into the Moderate class, ratios between 0.31 and 0.53 are labelled as Poor and any value less than or equal to 0.31 is considered Bad [46].

For example, a water sample collected at three depths—0 meters, 5 meters, and maximum depth—yielded Chl-a values of 5.20 μg/L, 4.94 μg/L, and 9.03 μg/L, respectively. Each of these values falls within the intermediate range (3–10 μg/L), and thus each received a score of three points. Dividing this score by the reference value (4.7) results in a score ratio of approximately 0.63 for all depths. According to the classification scheme, a score ratio of 0.63 corresponds to the Moderate ecological class [46]. This method ensures that classification is consistent and comparable across different depths and sampling periods, while remaining anchored to a defined ecological baseline.

For the second classification category, the focus shifts to physical-chemical determination rules that evaluate ecological status based on key water quality parameters [46]. These include DO, TN, TP, and water transparency. However, in this study, water transparency was not included because local authorities did not use it for classification purposes. Each parameter is assessed independently, and classification is assigned based on threshold values for three quality levels: Excellent, Good, and Moderate [46].

In the case of DO, a concentration between 8.8 mg/L and 11.6 mg/L is classified as Excellent. Values ranging from 5.0 mg/L to 8.8 mg/L, or above 11.6 mg/L, are considered Good, while any concentration equal to or below 5.0 mg/L is classified as Moderate [46]. For example, dissolved oxygen measurements taken at 0 metres, 2.5 metres, 5 metres, 10 metres, 15 metres, and the maximum depth ranged from 6.000 mg/L to 8.490 mg/L. All of these fall within the range of 5.0 to 8.8 mg/L, resulting in a classification of Good across all sampled depths [46].

Similarly, for TN, concentrations below 0.30 mg/L are rated as Excellent, values between 0.30 and 0.8 mg/L are considered Good, and concentrations greater than or equal to 0.8 mg/L fall into the Moderate category [46]. For TP, concentrations less than 0.04 mg/L are considered Excellent, those between 0.04 and 0.07 mg/L are rated as Good, and values equal to or exceeding 0.07 mg/L are classified as Moderate. These classification thresholds are applied uniformly across the depth profile to assess the ecological status of the water column using physical-chemical indicators [46].

For the last category, in particular, hydromorphological conditions were not verified due to insufficient sample collection. Therefore, aquatic life is assumed to be present but is not included in the final classification.

These rules need to be applied each month, and then a group-by of the year can be calculated. The datasets received from the local authorities do not include the WQI feature; they report water quality only on a yearly basis in their final report [46]. Therefore, it needs to be feature engineered to obtain the WQI necessary for supervised learning of the DL model [46].

The datasets used in this study originate from two distinct lakes in the Azores, Lake Fogo and Lake Azul, and include a wide range of meteorological and in-situ water-quality measurements. Each record consists of the year and month of collection. In-situ parameters are measured at multiple depths (0 m, 2.5 m, 5 m, 10 m, 15 m, and maximum depth), including Temperature (ʒC), DO (mg/L), PH, Conductivity (S/cm), Ortho-Phosphate (mgP-PO_4/L), TP (mgP/L), NO_3 (mgN-NO_3/L), TN (mgN/L), BOD (mgO_2/L), COD (mgO_2/L) and Chl-a (g/L), which can be seen on Table 3.1.

For Lake Fogo, the dataset contains only 13 monthly entries, with many months missing across the study period. All measured features are complete for the available months, with no partial missing values. Whereas Lake Azul, the dataset contains 48 monthly entries but exhibits scattered missing values across many parameters. Some features, such as Temperature, DO, and PH at deeper depths, have up to 13 missing entries, while others like BOD present larger gaps (over 20 missing entries in deeper measurements). In both datasets, the WQI target feature is absent, so it must be computed from the available parameters before any supervised modelling can be performed.

3.1.3 Weather Data Acquisition

The tabular dataset contains all the necessary components to identify the WQI of the water. Still, for DL models, it may be beneficial to access historical weather data to help them predict the target feature. Therefore, we gathered the coordinates of the two lakes, went to https://open-meteo.com/ [ref open-meteo], accessed historical weather data, and downloaded it to add to the tabular dataset. We gathered

Table 3.1: Features of the Weather Dataset

Feature	Description	Data Type
Ano	Year of measurement	Integer
Mês	Month of measurement	Integer
Temperatura	Water temperature at various depths (°C)	Float64
Oxigenio Dissolvido	Dissolved oxygen at various depths (mg/L)	Float64
pH	pH level at various depths	Float64
Conductividade	Water conductivity at various depths (S/cm)	Float64
Ortofosfatos	Orthophosphates at various depths (mgP-PO_4/L)	Float64
Fosforo Total	Total phosphorus at various depths (mgP/L)	Float64
Nitratos	Nitrates at various depths (mgN-NO_3/L)	Float64
Azoto Total	Total nitrogen at various depths (mgN/L)	Float64
Carencia Bioquimica Oxigenio	Biochemical Oxygen Deficiency at various depths (mgO_2/L)	Float64
Carencia Quimica Oxigenio	Chemical Oxygen Deficiency at various depths (mgO_2/L)	Float64
Clorofila a	Chlorophyll-a at various depths (g/L)	Float64

the ambient temperature, rainfall in millimetres, and short-wave radiation, as displayed in Table 3.2. These were the chosen parameters because the first can change the conditions of surface water, leading to increased algal blooms; the second can dilute several biological components in the water; and the last can also lead to algal blooms as they grow with sunlight radiation.

Table 3.2: Features of the Datasets for Lake Azul and Lake Fogo

Feature	Description	Data Type
Year	Year of measurement	Integer
Month	Month of measurement	Integer
Ambient Temperature	Ambient temperature in (°C)	Float64
Rain	Rainfall (mm)	Float64
Short-Wave Radiation	Short-Wave Radiation (MJ/m^2)	Float64

3.1.4 Satellite Data Acquisition

The satellite data is fetched from the Copernicus Application Programming Interface (API), which provides access to the Sentinel 2 mission, which began in 2015 and collects Earth imagery. The chosen bands were B01, B02, B03, B04, B05, B06, B07, B08, B8A, B09, B11, B12 and the Cloud Band (CLD), and the timeframe corresponds to exactly the tabular data spanning from 2017 to 2020, therefore obtaining exactly one set of images per month.

To obtain satellite images from the proper locations, a Region of Interest (ROI) needs to be delineated to identify the specific lake region for submission to the Copernicus API. This was done using Google Earth Engine to outline the lakes and export GeoJSON files containing the coordinates of the defined polygons for each lake.

The Copernicus API requires users to be authenticated on the platform; therefore, an account must be created, providing a private key to access the data. Once authenticated, the data retrieval process was fully automated using a Python-based workflow that communicates with the Copernicus API. This pipeline provides flexible, precise access to Sentinel-2 Level-2A surface reflectance products. The retrieval logic begins by defining the spatial and temporal query parameters: ROI, extracted from the exported GeoJSON polygons for each lake, and the date range spanning from January 2017 to December 2020. The workflow ensures spatial specificity by submitting the exact polygon geometry coordinates when querying the satellite database.

The system first queries the Copernicus catalogue to identify all available Sentinel-2 Level-2A products that intersect with the specified ROI and fall within the defined time range. Additionally, a cloud coverage filter is applied to limit image retrieval to scenes with acceptable atmospheric conditions. For each available acquisition date, a request is formulated that includes the selected spectral bands, the desired output resolution, and the specific time interval for which imagery is needed.

To ensure only the most suitable images are downloaded, the least cloud cover mosaicking strategy was employed, when retrieving the satellite images, only the ones that have the least clouds is taken. The cloud mask layer is retrieved alongside the optical bands to enable post-processing and the removal of cloudy pixels. Each successful request returns a compressed archive containing georeferenced Tagged Image File (TIF) images of the individual bands for a specific date and location. These files are saved in a structured directory format based on the lake's name with subdirectories for each year and month, making future data organisation and access straightforward.

As we are working on the Azores islands, even with the least-cloudy images, there are months when the ROI is either fully covered or partially cloudy. These clouds block the view of the lake region, and therefore, it is not possible to extract any information about the lake's water quality. Figure 3.1 represents a pipeline of how the satellite images are retrieved from the Copernicus API.

3.2 Data Visualisation

To better understand the data collected, a visual analysis will be performed to interpret the information in section **??**.

3.2.1 Lake Specificities

The lakes are situated on the island *São Miguel* in the Azores archipelago, as described in the subsection 3.1.1, which can be seen on Figure 3.2. Lake Azul is situated on the west of the island, while Lake Fogo is located in the centre of the island.

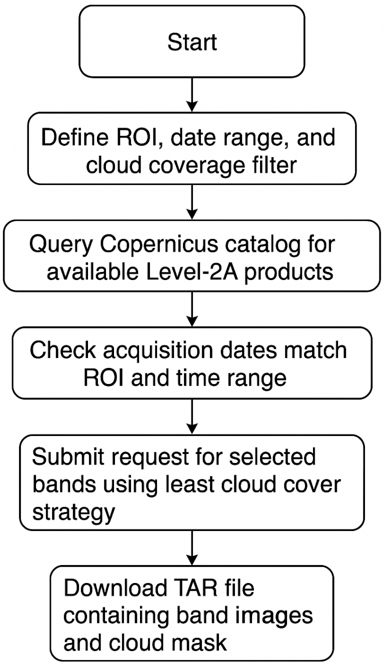

Figure 3.1: Pipeline for Satellite Image retrieval.

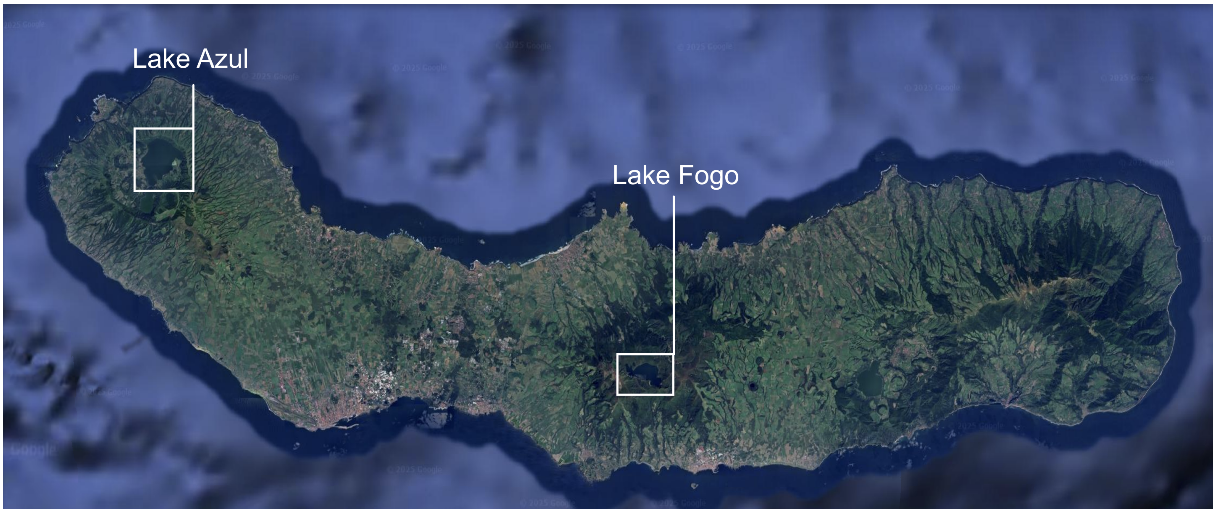

Figure 3.2: São Miguel Island Lake's Location. Provided by Copernicus Sentinel 2 [139].

A particular characteristic of Lake Azul is that it lies within a crater of an extinct volcano, making it deeper and requiring one to climb the crater to see it. Figure 3.3 makes it clear that when we observe around the lake, it is inside a similar style as a ravine.

As for Lake Fogo, considered the biggest lake of the Azores islands, which can be seen on Figure 3.4. This is again Lake Azul, an extinct volcano that Lake Fogo partially inundates. Hydrological analysis within the caldera reveals approximately 40 distinct watercourses, of which 18 contribute directly to the lake. The lake serves as a primary source for several downstream springs, which are crucial for the municipal water

Figure 3.3: Lake Azul. Provided by Copernicus Sentinel 2 [139].

supply of *Vila Franca do Campo*, *Lagoa*, *Ribeira Grande*, and *Ponta Delgada*.

Figure 3.4: Lake Fogo. Provided by Copernicus Sentinel 2 [139].

3.2.2 Spectrogram Analysis

A spectrogram analysis is performed to show the normalised NDCI values over 48 months, using Figure 3.5. The most notable feature is the clear, recurring pattern of dips and peaks. The dips occur roughly every

12 months (e.g., months 0, 12, 24, 36), indicating a strong seasonal cycle. These dips likely correspond to periods of low chlorophyll concentration, such as winter months, while the peaks correspond to periods of higher concentration, such as spring or summer.

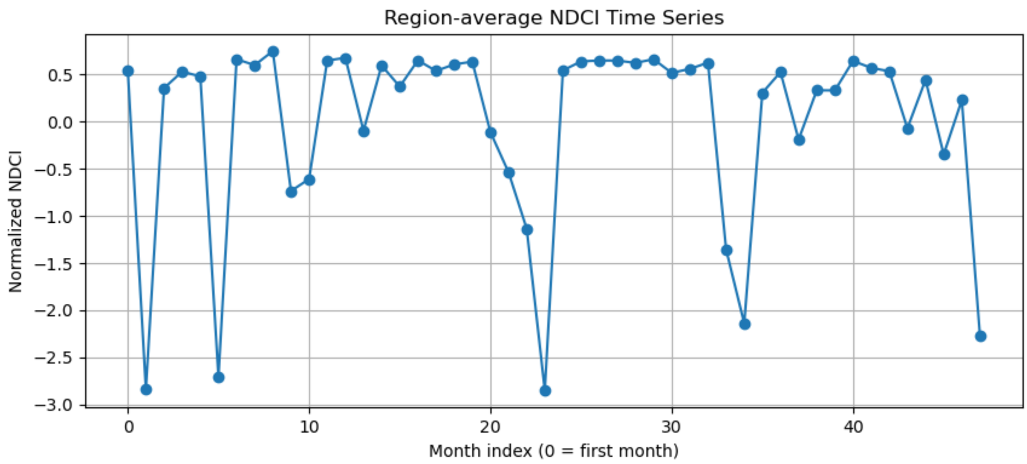

Figure 3.5: NDCI time-series for Lake Azul.

The Figure 3.6 plot shows the NDCI values over time. Unlike a smooth, regular seasonal pattern, this series is dominated by large, sharp drops in NDCI, particularly around month 20 and month 40. These significant, abrupt decreases suggest sporadic events that severely affect chlorophyll levels, such as extreme weather, volcanic activity, or sudden ecological shifts. The periods between these events show relatively stable NDCI values, without a precise, regular annual rhythm.

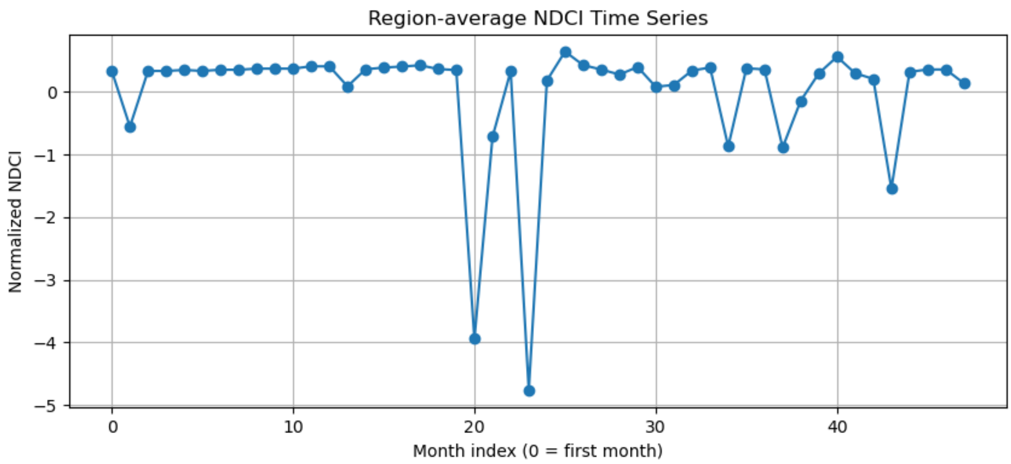

Figure 3.6: NDCI time-series for Lake Fogo.

The figure 3.7 image confirms the periodicity seen in the time series. The spectrogram uses a STFT to show how the frequency content of the NDCI signal changes over time. STFT is a technique used to analyse the frequency content of a signal as it changes over time [104]. Unlike the traditional Fourier Transform, which gives a single spectrum for the entire signal, STFT breaks the signal into smaller, overlapping segments and computes the Fourier Transform for each. This allows us to see how the signal's frequencies evolve. The following equation explains how this technique is used:

$$X(t,f) = \int_{-\infty}^{\infty} x(\tau)w(\tau - t)e^{-j2\pi f\tau}d\tau$$

where $X(t,f)$ is the output of the equation, representing the frequency content at time t and frequency f. The input signal is denoted by $x(\tau)$, and $w(\tau - t)$ is the window function, which is shifted by time t to select a short segment of the signal. The complex exponential, $e^{-j2\pi f\tau}$, is fundamental to the Fourier Transform and is used to analyse frequency components. In this context, j is the imaginary unit, defined as $\sqrt{-1}$, and τ is the time variable used for integration.

The bright horizontal band at approximately 0.083 cycles/month (1/12 cycles/month) indicates a strong, persistent signal at the annual frequency. The brightness of this band suggests that the yearly cycle is the dominant temporal pattern in the NDCI data throughout the entire time period analysed.

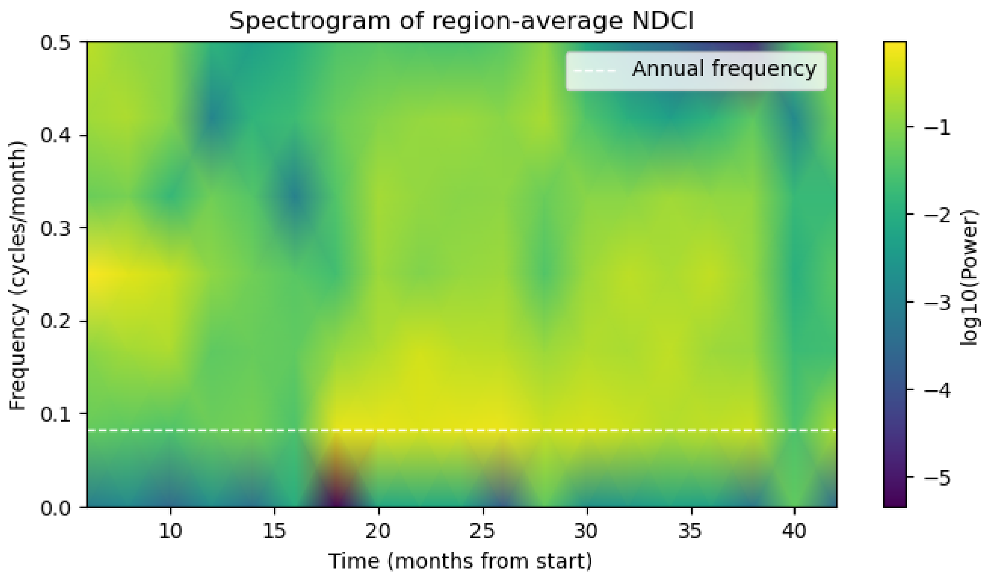

Figure 3.7: NDCI Spectrogram using STFT for Lake Azul.

A frequency analysis over time is provided by Figure 3.8. The horizontal white line marks the annual frequency (1/12≈0.083 cycles/month). Instead of a consistent bright band along this line, the spectrogram shows a strong, but intermittent, power signal at the annual frequency. This high-power signal is concentrated around the time periods where the sharp NDCI drops occurred, specifically between months 15-30 and again around months 35-45. This confirms that the significant drops are the primary drivers of the observed "annual"cycle power, rather than a steady, year-round seasonal fluctuation. The high power also extends to other frequencies during these event periods, suggesting the events are not perfectly periodic and contain a mix of frequency components.

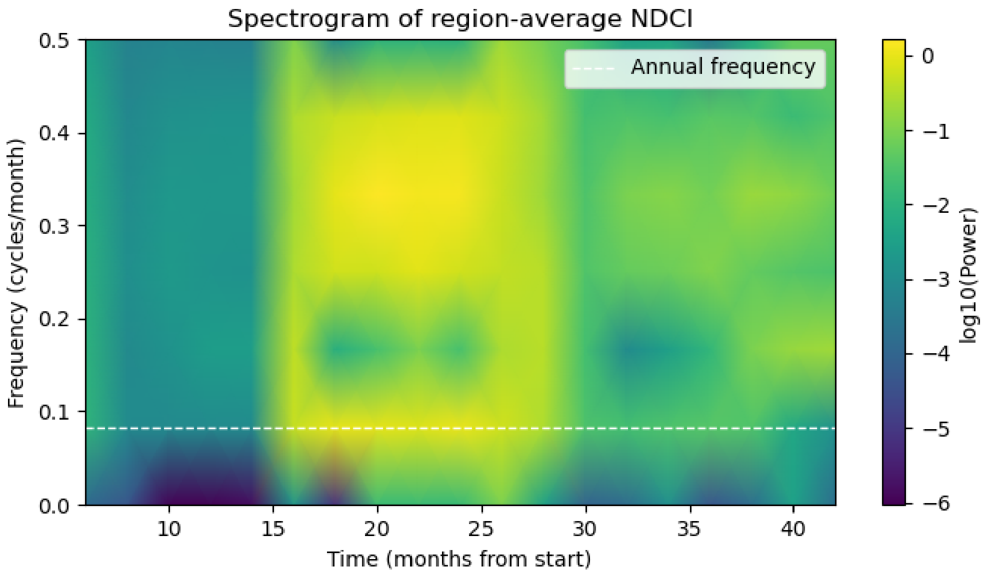

Figure 3.8: NDCI Spectrogram using STFT for Lake Fogo.

A spatial view of the annual cycle's strength across the region is provided through Figure 3.9. The map displays the power of the yearly frequency component for each pixel. Areas with higher annual power, represented by the brighter yellow-green colours, experience more pronounced seasonal changes in NDCI. Conversely, areas with lower power (darker purple colours) show less seasonal fluctuation. The map suggests that the most substantial seasonal variability is concentrated in a specific part of the region, namely the north-east and south section, possibly an area with more dynamic chlorophyll activity or one particularly affected by seasonal environmental factors.

The spatial distribution of the annual cycle's power is shown by 3.10. It reveals that the regions with the most significant "annual"signal (brighter yellow-green areas) are not uniform. Instead, the high-power areas are localised to specific spots, particularly in the south-east and north-west of the map. This suggests that the powerful, intermittent events seen in the time series and spectrogram are not affecting the entire region equally. Instead, their impact is spatially concentrated, highlighting specific areas that are more susceptible to these disturbances or where the disturbances originated.

3.3 Data Preparation

In this section, we will prepare the tabular and satellite data in order to address the issues presented in section **??**.

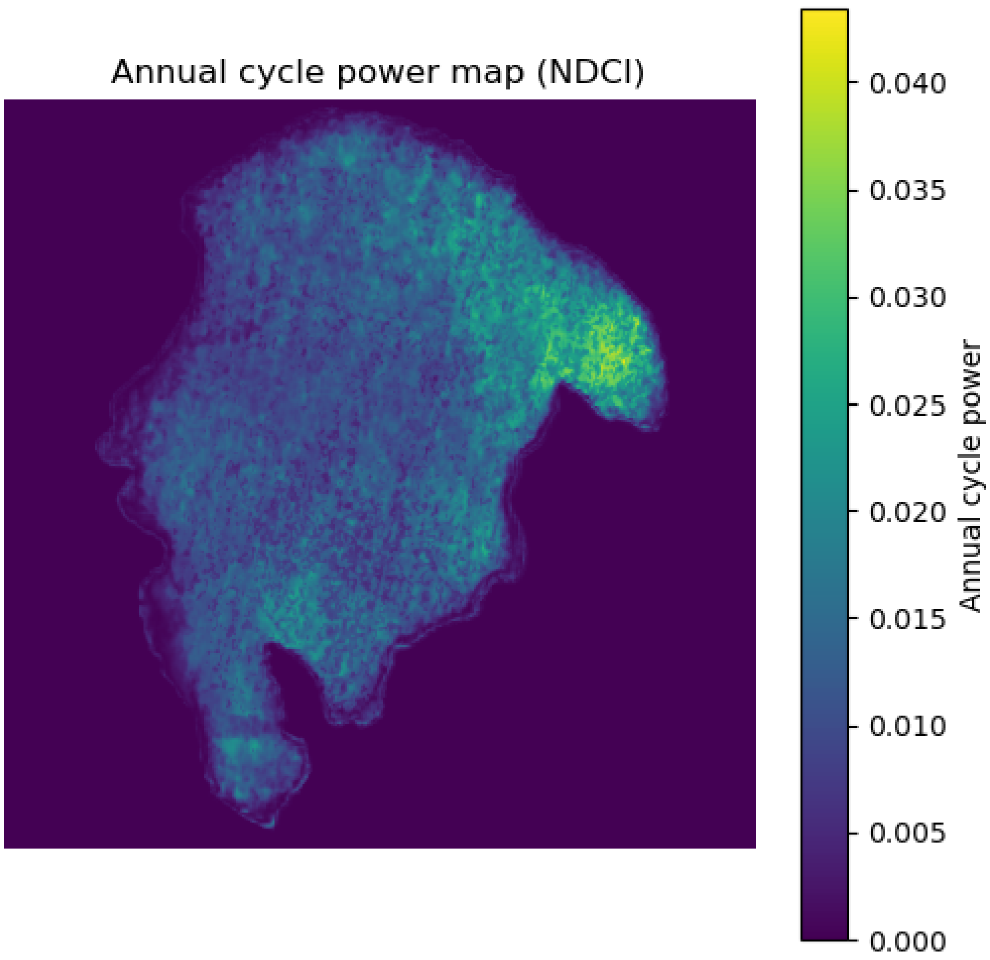

Figure 3.9: NDCI Annual Power Map for Lake Azul.

3.3.1 Tabular Data Preparation

For the tabular data, we first need to convert the data received in Excel for both lakes into separate sheets and parse it into a much simpler tabular format, for example Comma Separated Values (CSV). A simple script was used to open the relevant Excel files and convert each sheet to the aforementioned CSV format. Once this was accomplished, we needed to prepare the data in the correct format, as shown in Table 3.1. In particular, the date feature was only a simple string, so we engineered it into a datetime type and, for simplicity, split the year and month into separate columns. Following up on this, the data needed further preprocessing to fill in missing months, specifically for the Lake Azul dataset. For that, the moving average technique is employed to fill gaps using data from previous or subsequent years for the same month, to

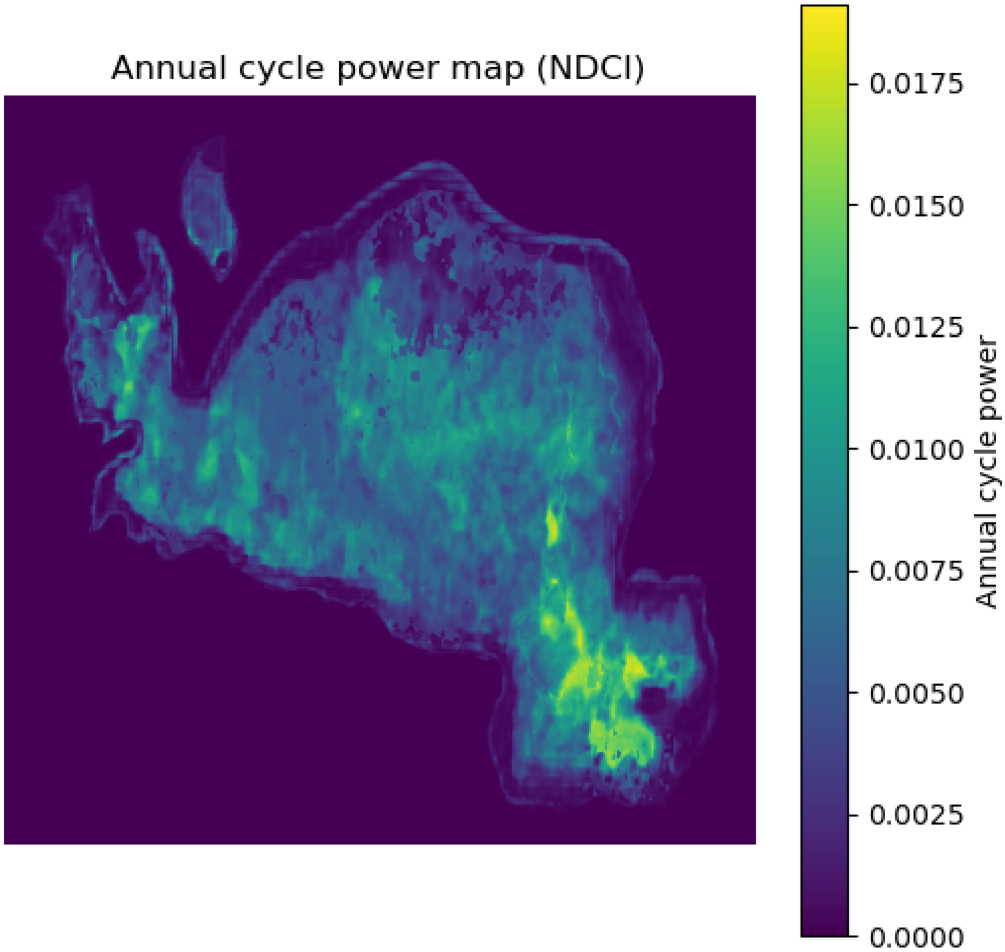

Figure 3.10: NDCI Annual Power Map for Lake Fogo.

remain consistent. The moving average can be explained in the following mathematical formula:

$$x_t = \begin{cases} x_t, & \text{if } x_t \text{ is observed,} \\ x_t = \frac{1}{n}\sum_{j=1}^{n} x_{t-12j} & \text{if } x_t \text{ is missing, where } x_{t-12j} \text{ is valid} \end{cases} \quad (3.1)$$

where x_t is the value at time t (with monthly sampling), n is the number of previous years considered in the calculation, and x_{t-12j} denotes the value from the same calendar month j years before time t. It allows us to average the values for the same month, only changing the year. The division by n ensures that the moving average x_t represents the mean of the values from the same month in the past n years. As for the Lake Fogo dataset, the necessary preprocessing was completely different; this dataset only had entries every 3 months, so the months in between needed to be filled in. For that reason, a linear interpolation was implemented, which is a widely used technique for estimating missing values in time series and spatial datasets due to its simplicity, efficiency, and minimal assumptions about the underlying data distribution. Connecting two known data points with a straight line and estimating intermediate values along this line preserves the local trend of the data without introducing abrupt discontinuities. A mathematical formula for better clarification is provided as follows:

$$x_{(y,m)} = x_{(y_0,m_0)} + \frac{x_{(y_1,m_1)} - x_{(y_0,m_0)}}{\Delta M} \cdot \Delta m \tag{3.2}$$

Where $x_{(y,m)}$ is the interpolated value for the target year y and month m, $x_{(y_0,m_0)}$ is the value from the closest available date before (y,m), and $x_{(y_1,m_1)}$ is the value from the closest available date after (y,m). ΔM represents the total number of months between (y_0,m_0) and (y_1,m_1), while Δm denotes the number of months from (y_0,m_0) to the target date (y,m). This interpolation is applied independently to each column (parameter) in the dataset, ensuring that missing entries are estimated solely from the nearest available time values. For both datasets, additional feature engineering was needed because the target feature, namely WQI, was not available for the DL models to use. Consequently, the rules explained in section 3.1.2 needed to be applied for all entries in the datasets to create the correct WQI feature. To comply with the local authorities' rules, the following steps were taken. The Chl-a concentration, denoted as C_a, is first converted into an intermediate score S_a. This score is then normalised and used to determine the final classification, Chl-a class, which ranges from 0 (Excellent) to 4 (Bad).

$$S_a = \begin{cases} 5, & \text{if } C_a < 3 \\ 3, & \text{if } 3 \leq C_a < 10 \\ 1, & \text{if } C_a \geq 10 \end{cases} \tag{3.3}$$

$$\mathit{Chl\text{-}a}_{class} = \begin{cases} 0, & \text{if } S_a/4.7 > 0.94 \\ 1, & \text{if } S_a/4.7 > 0.74 \\ 2, & \text{if } S_a/4.7 > 0.53 \\ 3, & \text{if } S_a/4.7 > 0.31 \\ 4, & \text{otherwise} \end{cases} \tag{3.4}$$

The DO concentration, C_{do}, is classified into three categories based on predefined ranges. A classification of 0 (Excellent) is assigned for values within an optimal range, 1 (Good) for values slightly outside this range, and 2 (Moderate) for all other values.

$$DO_{class} = \begin{cases} 0, & \text{if } 8.8 < C_{do} < 11.6 \\ 1, & \text{if } (5.0 < C_{do} \leq 8.8) \lor (C_{do} \geq 11.6) \\ 2, & \text{otherwise} \end{cases} \tag{3.5}$$

The TN concentration, C_{tn}, is classified into three categories based on a simple threshold method.

$$TN_{class} = \begin{cases} 0, & \text{if } C_{tn} < 0.30 \\ 1, & \text{if } 0.30 \leq C_{tn} < 0.8 \\ 2, & \text{otherwise} \end{cases} \tag{3.6}$$

The TP concentration, C_{tp}, is classified similarly to TN using concentration thresholds to determine its category.

$$TP_{class} = \begin{cases} 0, & \text{if } C_{tp} < 0.04 \\ 1, & \text{if } 0.04 \leq C_{tp} < 0.07 \\ 2, & \text{otherwise} \end{cases} \tag{3.7}$$

The final WQI for a given time point is determined by the maximum value of the four individual parameter classifications, reflecting the "worst-case"scenario among the measured parameters.

$$WQI = \max(Chl\text{-}a_{class}, DO_{class}, TN_{class}, TP_{class}) \tag{3.8}$$

At the end, we have the weather data that needs to be merged with the lake's dataset. This weather dataset also required preprocessing by engineering the date into the correct datetime type and separating it into year and month features. Then, the merging was relatively simple because of the preprocessing done to both lakes' datasets of the date feature, by merging on the year and month features.

For a better understanding of the lake's WQI feature, a histogram represents how many of each entry have been created, which can be observed on Figures 3.11 and 3.12. The first one represents Lake Azul, while the latter one represents Lake Fogo.

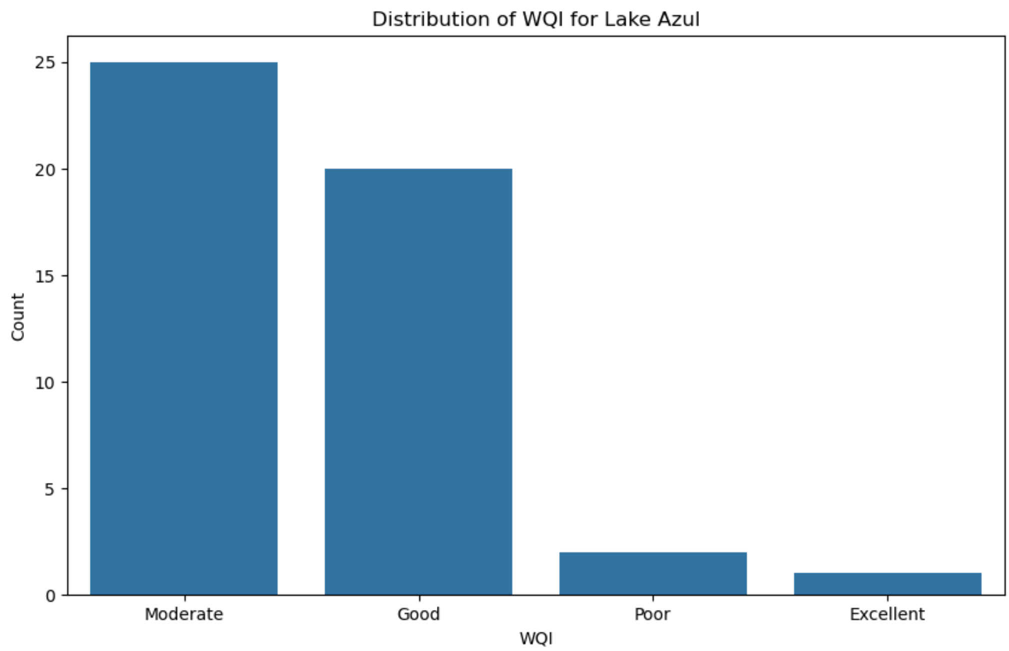

Figure 3.11: WQI distribution of Lake Azul.

For Lake Azul, there are four possible water quality classes, with the majority falling into Good or Moderate, indicating good overall quality. Moderate has 25 entries, while Good has 20 entries. There have been 2 entries classified as Poor and 1 entry classified as Excellent. As one can see, Lake Azul's water quality has never been Bad for the years between 2017 and 2020.

Regarding Lake Fogo, again, the majority of the classifications are either Good or Moderate, but the Good class has more entries, 28, to be precise. The Moderate one has eleven entries. Additionally, the WQI has been classified eight times as Excellent and once as Poor, indicating that the quality has never dropped to Bad during the observed timeline.

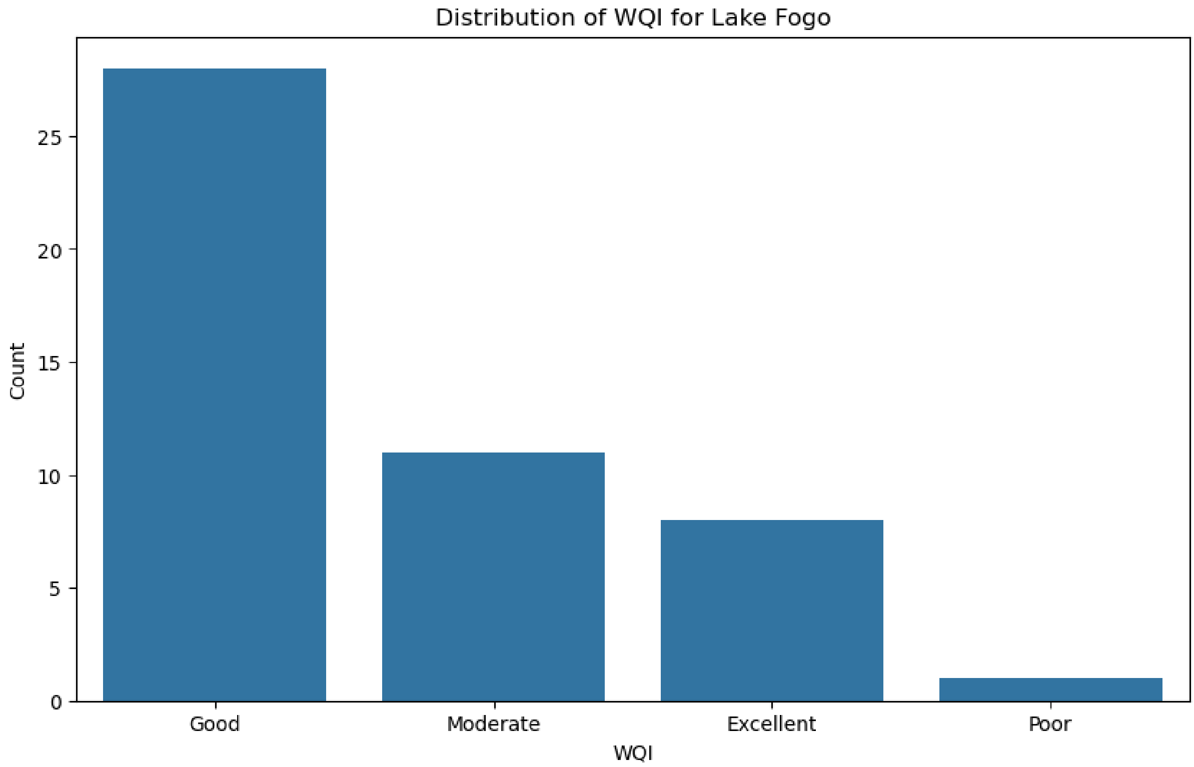

Figure 3.12: WQI distribution of Lake Fogo.

3.3.2 Satellite Data Preparation

The satellite data has an entirely different preprocessing step than its tabular data counterparts. Here, we are dealing primarily with clouds. The approach taken was either to remove the clouds by turning these pixels black or, if the cloud intensity was really high, to apply a threshold-based scheme similar to the moving average. Although this time to reconstruct the image was completely sufficient, it still required more than simply going back in time to get the average from previous years. The priority for reconstructing the images was to gather all images from the current month and select only those that were not obscured by clouds, then "paste"them together to create a clear region of the lake. If this was not possible, we took the same month from the previous years and calculated the average to create the new image. If the previous year's same month was also unusable due to cloud cover, we calculated the average of the six previous months. This needed to be done because of the very cloudy region. The Azores islands are well known for their very cloudy conditions, which make it a very challenging environment for RS using satellite imagery. The previously explained steps are backed by the following equations to better showcase the procedure.

First, a cloud mask is applied. For a pixel at location (i,j) at time t, if the cloud intensity C_t(i,j) is below a certain threshold (80% in our case), the pixel is set to zero (black). An 80% threshold is a conservative choice that prioritises preserving valid, bright ground data over removing every faint, semi-transparent cloud. The value of the pixel, P_t(i,j), is consequently defined as:

$$P_t(i,j) = \begin{cases} V_t(i,j), & \text{if } C_t(i,j) \leq 80\% \\ \text{Missing}, & \text{if } C_t(i,j) > 80\% \end{cases} \tag{3.9}$$

where $V_t(i,j)$ is the original pixel value and "Missing"indicates that the pixel needs to be reconstructed.

$\hat{P}_t(i,j)$, follows a tiered approach based on data availability, by either using data from the same current month. If a pixel is missing at time t (e.g., in a specific year and month), the first step is to check for valid pixels from other images of the same month and year that were not obscured by clouds. The reconstructed pixel value is then a composite of these valid values. Let S_t be the set of all images for month t. The reconstructed pixel is:

$$\hat{P}_t(i,j) = P_{t'}(i,j), \quad \text{for some } t' \in S_t \text{ where } P_{t'}(i,j) \text{ is not missing} \tag{3.10}$$

The next approach, if the previous one does not work, is to use data from previous years. If a complete image cannot be formed using data from the same month and year, the system averages the corresponding pixels from the same month in previous years. Let t_m be the current month. The reconstructed pixel is:

$$\hat{P}_t(i,j) = \frac{1}{k} \sum_{y=1}^{k} P_{(t_m - 12y)}(i,j) \tag{3.11}$$

where $P_{(t_m-12y)}(i,j)$ is the pixel value for the same month in the y-th previous year, and k is the number of valid years available for the average.

As for the last approach, the usage of the previous six months' data. If the data from the same month in previous years is also missing or not usable, the method falls back to a six-month moving average of the most recent valid data, which can be described using the following mathematical expression:

$$\hat{P}_t(i,j) = \frac{1}{6} \sum_{i=1}^{6} P_{t-i}(i,j) \tag{3.12}$$

This equation calculates the average of the six previous months' pixel values to fill in the missing data for the current month.

When the images are in a usable state, meaning they have no clouds, indexes can be applied across the different bands to produce more information-rich images. The indexes applied are the following NDWI, NDCI, FAI, Harmful Algal Blooms (HABs), Coloured Dissolved Organic Matter (CDOM), turbidity, water clarity, and TrueColour. These indices can be derived by combining satellite bands, as with the NDWI, which is used to delineate open water bodies and is calculated from reflectance in the green and NIR bands. The formula is a normalised ratio that measures the difference between the two spectral bands, with the result then clipped to the range -1 to 1. A higher NDWI value generally indicates a greater presence of water.

$$NDWI = \frac{R_{\text{green}} - R_{\text{nir}}}{R_{\text{green}} + R_{\text{nir}}} \tag{3.13}$$

This index was mainly used to crop the lake, so that the following indices are applied solely to the water body itself, excluding potential unwanted areas.

The following index is related to Chl-a concentration, likely the NDCI, and is calculated from the red-edge and red bands. The formula is a normalised ratio that highlights Chl-a content in water bodies, with the result clipped to a range of -1 to 1. Higher values of this index typically correlate with higher concentrations of Chl-a.

$$NDCI = \frac{R_{\text{red\_edge}} - R_{\text{red}}}{R_{\text{red\_edge}} + R_{\text{red}}} \tag{3.14}$$

The Turbidity index uses the red and green spectral bands to estimate the amount of suspended sediment or other matter in the water. The formula measures the difference between red and green reflectance, then clips the result to the range -1 to 1. Higher turbidity values indicate more suspended particles and therefore less clear water.

$$Turbidity = \frac{R_{\text{red}} - R_{\text{green}}}{R_{\text{red}} + R_{\text{green}}} \tag{3.15}$$

The CDOM is used to estimate the concentration of CDOM in water, which is a key indicator of water quality. The formula calculates a normalised difference using the aerosol and blue bands, with the result clipped to a range of -1 to 1. Higher CDOM values suggest a greater presence of organic matter that absorbs light in the blue and ultraviolet parts of the spectrum.

$$CDOM = \frac{R_{\text{aerosol}} - R_{\text{blue}}}{R_{\text{aerosol}} + R_{\text{blue}}} \tag{3.16}$$

The Water Clarity index is calculated as a simple ratio of the blue and green spectral bands. The formula, which is also clipped to the range -1 to 1, is a simple metric that generally indicates clearer water with less suspended material that would otherwise absorb and scatter light.

$$WaterClarity = \frac{R_{\text{blue}}}{R_{\text{green}}} \tag{3.17}$$

The HABs index is a specific formula designed to detect the presence of harmful algal blooms by combining information from the blue, red, and NIR bands. The formula is a linear combination of these bands that is sensitive to the spectral signatures of certain types of algae, with the result clipped to the range -1 to 1.

$$HABs = R_{\text{nir}} - (R_{\text{red}} + 0.5 \times (R_{\text{nir}} - R_{\text{blue}})) \tag{3.18}$$

For the last index, FAI is an advanced index used to detect and quantify floating algal mats and other surface scums. The formula performs a baseline subtraction to isolate the spectral signature of floating algae from the background water. It uses reflectance values from the NIR, red, and SWIR bands, to correct for atmospheric and water-body effects.

$$FAI = (R_{\text{nir}} - R_{\text{red}}) - (R_{\text{swir}} - R_{\text{red}}) \times \frac{\lambda_{\text{nir}} - \lambda_{\text{red}}}{\lambda_{\text{swir}} - \lambda_{\text{red}}} \tag{3.19}$$

The satellite data still has a final preprocessing step: augmenting the images to increase the number of them. These augmentation techniques can be applied using horizontal flipping, affine transformations (such as rotation, shear, uniform scaling, and translation), and brightness and contrast adjustment. These can be applied to RGB and greyscale images, but saturation and hue adjustments can only be used to RGB images. Image augmentation techniques can be expressed mathematically to clarify how they modify pixel values and spatial structures. Below, we outline some common transformations and their underlying equations. First, horizontal flipping can be written as:

$$I'(x, y, c) = I(W - 1 - x,\ y,\ c) \tag{3.20}$$

where, W is the image width, x and y are pixel coordinates, and c is the colour channel index. This operation mirrors the image across its vertical axis, making the model invariant to left–right orientation changes.

Beyond flipping, we often need to simulate more complex geometric transformations. These include rotation, shear, scaling, and translation, which can be combined into a single affine transform:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = s \cdot \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \cdot \begin{bmatrix} 1 & \tan\phi \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{3.21}$$

This affine transform applies a rotation by θ, a shear by ϕ, a uniform scaling by s, and a translation by (t_x, t_y), simulating changes in orientation, perspective, object size, and position.

In addition to geometry, photometric augmentations modify pixel intensities. A simple yet effective one is brightness adjustment:

$$I'(x, y, c) = I(x, y, c) \cdot \beta \tag{3.22}$$

The parameter $\beta \in [1 - b,\ 1 + b]$ controls brightness. A value $\beta > 1$ brightens the image, while $\beta < 1$ darkens it, simulating different lighting conditions.

Another important transformation is contrast adjustment, which changes the spread of pixel values relative to the mean:

$$I'(x, y, c) = \alpha \cdot \big(I(x, y, c) - \mu\big) + \mu \tag{3.23}$$

Here, μ is the mean pixel intensity across the image, and $\alpha \in [1 - c,\ 1 + c]$ is the contrast factor. Higher α increases intensity differences from the mean, making the image appear sharper; lower α reduces contrast.

When working in HSV colour space, we can further manipulate saturation, which controls colour vividness:

$$S' = \alpha_s \cdot S \tag{3.24}$$

where S is the saturation channel in Hue-Saturation Value (HSV) colour space, and $\alpha_s \in [1-s,\ 1+s]$ adjusts how vivid or dull the colours are, which cannot be applied to greyscale images.

Finally, colour shifts can be simulated by modifying the hue channel in HSV space:

$$H' = (H + \Delta_h) \bmod 1 \tag{3.25}$$

Here, H is the hue channel in HSV space, and $\Delta_h \in [-h,\ h]$ shifts the colour tone of the entire image. Hue adjustment helps the model generalise to variations in object colour, but it is not applied to greyscale images.

The Figures 3.13 and 3.14 represent the NDCI index, but these images a colour ramp has been applied. The areas with a brighter, redder colour represent those with a higher concentration of NDCI values. For Lake Azul, these spots are concentrated mainly on the west and northeast sides, where small red spots appear.

Figure 3.13: Colour Ramp for NDCI of Lake Azul.

Figure 3.14: Colour Ramp for NDCI of Lake Fogo.

In the Figure for Lake Fogo, one can see that the red spots mainly appear on the west side, highlighting the higher NDCI values. These images, created using the indices, provide us and the DL models with more knowledge to better understand the underlying properties of each water body's parameters.

3.4 Summary

Chapter 3 outlined the materials, data sources, and methodological steps employed in this research. The study focused on two Azorean Lakes, specifically Lake Azul and Lake Fogo, selected for their ecological importance and existing water-quality monitoring initiatives. The chapter detailed the acquisition of three main datasets: in-situ tabular data (2017–2020) containing physicochemical and biological water parameters; meteorological data to account for climatic influences; and Sentinel-2 satellite imagery retrieved via the Copernicus API. The integration of these datasets established a comprehensive foundation for linking RS observations to ground-truth water quality indicators.

Subsequent sections focused on data visualisation and preparation. Visual analyses, including spectrograms and annual power maps, revealed temporal and spatial patterns of chlorophyll concentration

(NDCI), indicating seasonal fluctuations and event-driven changes. Preprocessing steps involved formatting tabular data, handling missing values using moving averages, and organising satellite imagery for model input. Image augmentations, such as brightness, contrast, and hue adjustments, were applied to improve model generalisation. Collectively, this chapter established the methodological groundwork necessary for the experimental framework described in the following chapter.

Chapter 4

Experiments

The following chapter will mainly focus on the experimental setup of this dissertation. In section 4.1, the different configurations used to perform the study are described. Another essential section 4.2 describes the architectural setup of the DL models used. Section 4.3 explains the metrics used to evaluate the models' performance. At the end, section 4.4 explains the hardware and software used throughout this study.

4.1 Experimental Setup

The experimental setup of this study includes several approaches, not only training the data on different DL models but also transforming the data into different perspectives. By 'perspectives', we mean to cut around the in-situ collection point and keep the entire lake in view.

This decision was made based on the idea that the tabular data correspond to a sample site; therefore, it would be interesting to analyse only that specific point and the whole lake. Additionally, we trained the model on non-augmented and augmented data. We used pre-trained models to better understand whether data augmentation really helps the model and whether pre-trained models help reduce time while maintaining good results. It is also crucial to find the best setup for the DL models described in section 4.2.

The setup requires searching for the best hyperparameters, which we do using Grid Search in our case. Grid search is a simple algorithm that exhaustively searches a predefined subset of a model's hyperparameter space. For each experiment, the results are saved, including the hyperparameters and all metrics, for later analysis. Table 4.1 highlights all the hyperparameters used for each model employed. The batch size used is the same across all models. Swin's choice of image size of 24 instead of 25 is due to the requirement that image size divided by patch size must be even, to maintain symmetry in window

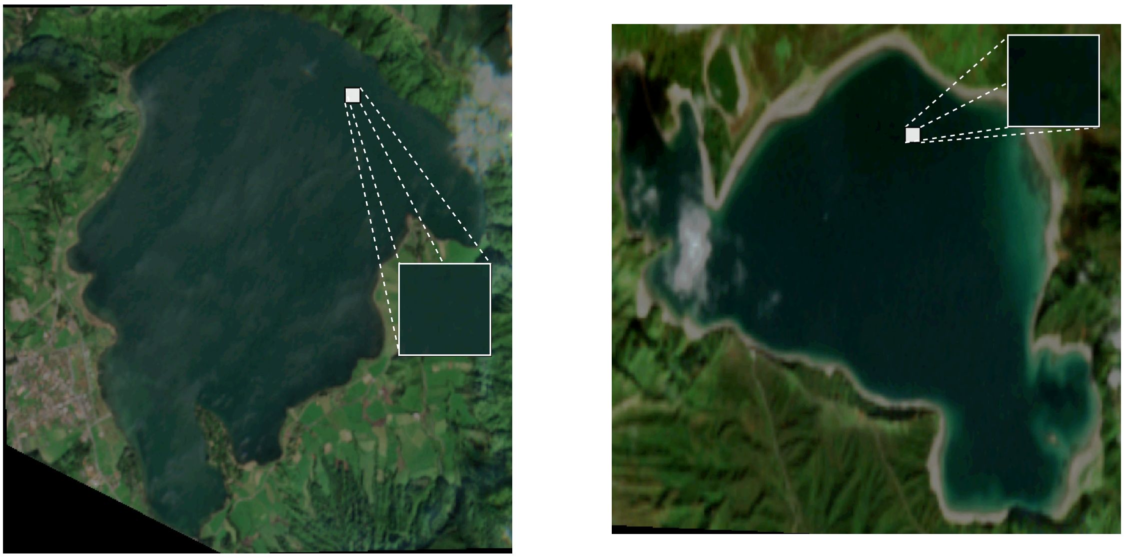

Figure 4.1: Crop of In-Situ Observation Site. Left Image: Lake Azul; Right Image: Lake Fogo

partitioning. The choice of small batch sizes is driven by the need for training stability on smaller datasets, especially for transformers, which can overfit quickly with large batches. DenseNet uses growth rates and BN sizes to control network width and bottleneck compression. Smaller growth rates save memory, larger values increase representational capacity.

For ResNet, no growth rate or embedding dimension is used, as it employs a simpler design with variable dropout rates instead. UNet employs base channels to balance between finer classification details and computational cost. ViT and Swin are very similar and therefore use many common hyperparameters, such as patch size, embed dim, depth, num heads, and Multilayer Perceptron (MLP). Ratio gives fine control over transformer complexity. On the other hand, Swin additionally uses window size and attention dropout for local attention and better regularisation. Dropout Rate is present in ResNet, UNet, ViT, Swin, where it ranges from no dropout to moderate regularisation (0.5). Additionally, Swin uses attention dropout with small values to prevent over-regularising attention mechanisms. Batch norm is deactivated across all models due to small batch sizes, which can make it very unstable. For L1 and L2 regularisation, weight decay and λ_{l1} are used with fixed small values to regularise the optimiser.

Another critical aspect of the experiments is that we employed cross-validation using stratified K-folds. It is a variant of K-Fold cross-validation that ensures each fold has approximately the same class distribution as the entire dataset, especially when the dataset is imbalanced. In a classification problem, regular K-Fold splitting might produce folds where some classes are overrepresented or even missing entirely. The stratified version tries to avoid this, but it does not entirely avoid it. If the dataset is small, it may struggle to include folds with all available classes. For future reproducibility, a fixed seed of 55 was employed across all experiments.

In a classification task, an optimiser is the component that updates your model's parameters to minimise the chosen loss function. It does so by adjusting the weights and biases in the direction that reduces

Table 4.1: Hyperparameter Search Spaces Specific to Each Architecture.

Parameter	DenseNet	ResNet	UNet	ViT	Swin Transformer
Image Size	[25, 128]	[25, 128]	[25, 128]	[25, 128]	[24, 128]
Batch Size	[2, 4, 8]	[2, 4, 8]	[2, 4, 8]	[2, 4, 8]	[2, 4, 8]
Growth Rate	[24, 32]	-	-	-	-
BN Size	[3, 4]	-	-	-	-
Base Channels	-	-	[32, 64]	-	-
Bias	[False]	[False]	-	-	-
Dropout Rate	-	[0, 0.5]	[0, 0.5]	[0, 0.5]	[0, 0.5]
Attention Dropout	-	-	-	-	[0.0, 0.1]
Activation Function	[relu, softmax]	[relu, softmax]	[relu, softmax, tanh, sigmoid]	[relu, gelu]	[relu, gelu]
Pooling	[max, avg]	[avg]	-	-	-
Batch Norm	[False]	[False]	[False]	-	-
Patch Size	-	-	-	[5, 16]	[6, 16]
Embed Dim	-	-	-	[32, 64]	[64, 96]
Depth	-	-	-	[2, 4]	-
Num Heads	-	-	-	[4, 8]	-
Window Size	-	-	-	-	[4, 8]
MLP Ratio	-	-	-	[2.0, 4.0]	[2.0, 4.0]
Learning Rate	[0.001, 0.0001]	[0.001, 0.0001]	[0.001, 0.0001]	[0.001, 0.0001]	[0.001, 0.0001]
Weight Decay	[0.0001]	[0.0001]	[0.0001]	[0.0001]	[0.0001]
λ_{l1}	[0.0001]	[0.0001]	[0.0001]	[0.0001]	[0.0001]

the error between predictions and ground truth. It uses the gradients computed in backpropagation to update parameters, controls how fast and in what manner the model learns and balances convergence speed and stability. In our case, we employed two types of optimisers, both based on Adam, but with different pre-training setups. For the pre-trained models, we use separate parameter groups, with the pre-trained backbones having lower learning rates to fine-tune slowly and preserve learned features. On the other hand, the specialised Layers and classification heads make use of a higher learning rate, so they adapt quickly to our dataset. For regular model training, a single learning rate is used across all layers, since there's no pretrained knowledge to preserve, which speeds up training in the early epochs. L2 regularisation is also used in the optimiser, penalising large weights. It keeps decision boundaries more straightforward and more generalisable by preventing overfitting.

Another aspect used in the experiments was callbacks. These are pieces of code that monitor the validation loss during training and perform certain actions when specific conditions are met. Instead of running a fixed training plan, callbacks help the model react to its own performance. In our case, we used Early Stopping and ReduceLROnPlateau. The first one's purpose is to stop training when no improvement is observed based on a specific metric over a certain number of epochs, also called patience. Key advantages include preventing overfitting and saving time and resources by avoiding unnecessary epochs. The second callback lowers the learning rate when the model stops improving, allowing finer adjustments to the weights. Its advantages are to help escape plateaus in optimisation, to avoid overshooting the minimum loss by using smaller learning rates later in training and to reduce the need to guess an optimal learning rate beforehand.

An additional essential aspect before starting to tune the models is to determine how many epochs are needed to tune them. Therefore, we created, for each model, eight learning curves, where the number of epochs is plotted on the x-axis and loss and accuracy on the y-axis. The high number of extracted learning curves is due to the fact that we generate them based on four different scenarios. The first one, illustrated in Figure 4.2, creates them using the first hyperparameters with 15 epochs, then, with the same hyperparameters, trains them for 100 epochs, as shown in Figure 4.3. These can be considered the more 'lighter' approach.

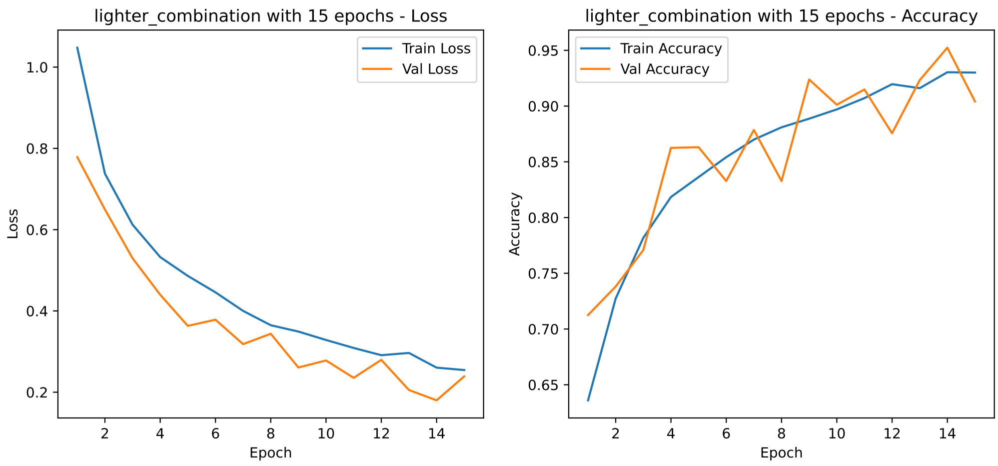

Figure 4.2: Learning Curve of Lighter Version on 15 epochs for DenseNet40 for Lake Fogo.

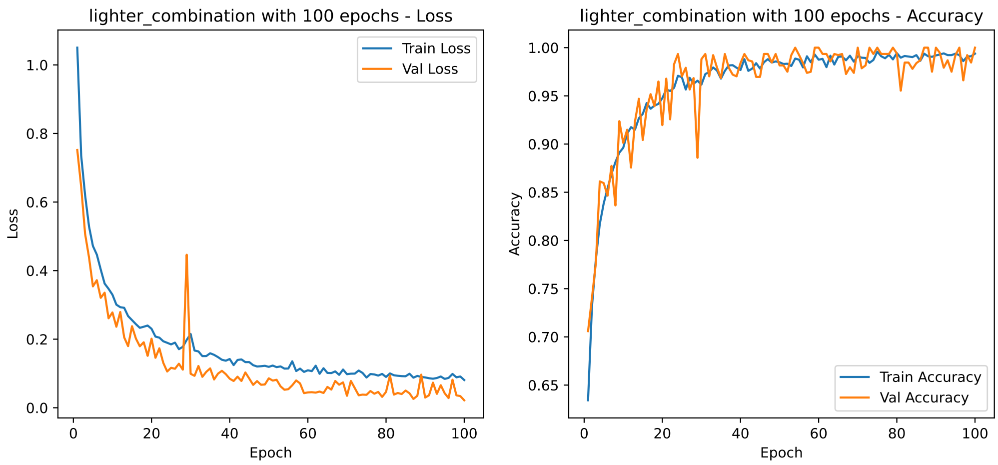

Figure 4.3: Learning Curve of Lighter Version on 100 epochs for DenseNet40 for Lake Fogo.

For the last two, namely Figures 4.4 and 4.5, the same is done as the first two, but the chosen hyperparameters are the last ones; these reflect the more 'heavier' approach. Afterwards, to select the correct number of epochs, we analyse each learning curve and choose based on stability between the training and validation curves, as well as where the accuracy peaks and/or the loss bottoms.

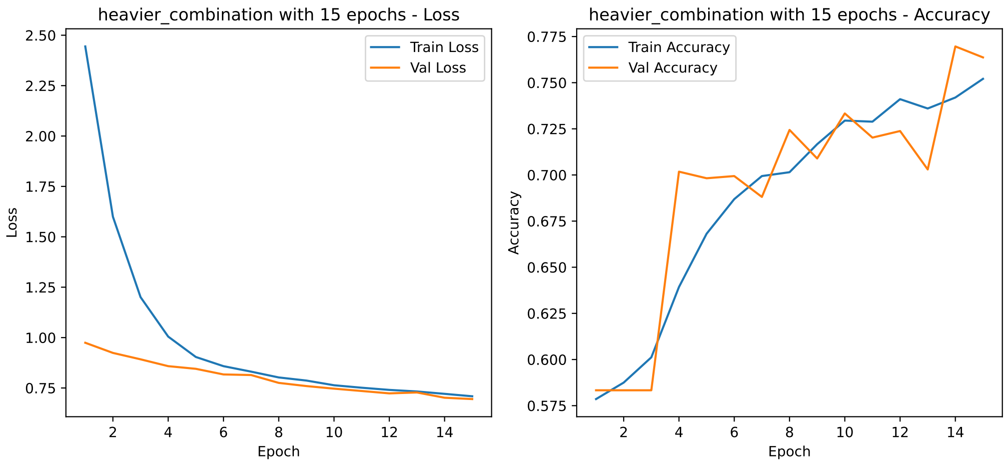

Figure 4.4: Learning Curve of Heavier Version on 15 epochs for DenseNet40 for Lake Fogo.

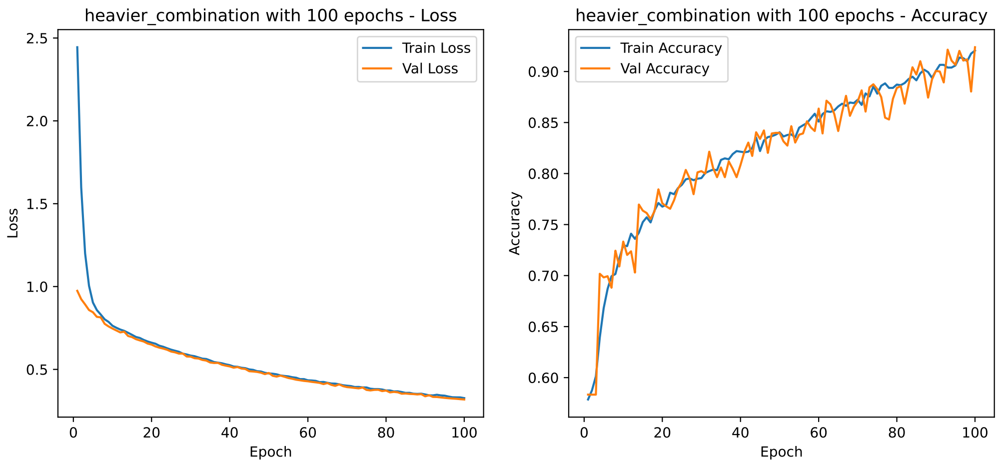

Figure 4.5: Learning Curve of Heavier Version on 100 epochs for DenseNet40 for Lake Fogo.

4.2 Deep Learning Model Conception

In this section, we will focus on the DL models we use, how they are constructed, and the pre-trained models from torchgeo [132] and Big Earth Net [21, 51]. Additionally, the custom implementation needs to handle different channel counts, especially for RGB and greyscale.

4.2.1 General Architectural Setup

The architectural setup for each DL model underneath shares common traits.

To accommodate heterogeneous inputs, the forward method iterates over all images in the batch. Each image is first expanded with an additional batch dimension to ensure compatibility with convolutional operations. Since the dataset contains both RGB and greyscale imagery, the model includes separate initial convolutional pathways for each. This design ensures that the distinct characteristics of each image type are preserved during feature extraction. Following the modality-specific convolutions, the extracted features are passed through a set of shared initial layers comprising batch normalisation, a non-linear activation, and pooling. These operations normalise the feature distribution, introduce nonlinearity, and reduce spatial resolution, preparing the feature maps for deeper processing. The outputs are then fed into the leading backbone network specific to each model.

At the end of the backbone, each image representation is flattened into a fixed-size one-dimensional vector. These per-image feature vectors are collected into a list and concatenated along the batch dimension, yielding a unified feature tensor. This tensor serves as the input to the classification stage.

The classification head is divided into 2 parts; the first uses a simple dense layer to process the features from the backbone. Meanwhile, the other strategy relies on concatenating tabular data, if this is available. It concatenates the image feature vector with the tabular input vector, then passes the combined features through a linear layer for final classification. The forward method automatically selects the classifier based on whether tabular data is provided. This hybrid approach enables the model to leverage both visual and non-visual information to improve classification performance.

4.2.2 DenseNet

Our DenseNet is a custom architecture designed to handle both image and optional tabular data for classification tasks. The core principles of the original DenseNet architecture are maintained, including dense blocks with feature concatenation and transition layers for downsampling, but with a flexible design that accommodates different input types and configurable parameters. The core of the DenseNet architecture is maintained by accommodating the dense layer, the dense block and the transition layer. The first one consists of a 1x1 convolution followed by a 3x3 convolution. Crucially, the input feature map is concatenated with the layer's output along the channel dimension. This feature reuse mechanism is what gives DenseNet its name and allows for the propagation of information and gradients throughout the network, mitigating the vanishing gradient problem. The second one is simply a sequence of multiple dense layer modules. Within a dense block, the number of feature channels grows with each layer. The output of each layer is concatenated with the input to the next layer, meaning the input to layer L consists of the feature maps from all preceding layers within the block. The last core concept is placed between dense blocks to perform downsampling. It consists of a 1x1 convolution to reduce the number of feature channels (a compression step) followed by a pooling operation to reduce the spatial dimensions of the feature map.

In the final solution, the network begins with a 7x7 convolution followed by batch normalisation, an activation function, and a pooling layer. A key feature of this implementation is its ability to handle both RGB and greyscale images by using separate initial convolutional layers. The forward pass dynamically selects the appropriate layer based on the image type provided. The central feature extraction part of the network is a sequential stack of dense blocks and transition layer modules. The configuration of these blocks is determined by the depth parameter, which corresponds to well-known DenseNet configurations, such as DenseNet-121. The backbone's output is passed through a final batch normalisation layer and a pooling operation.

4.2.3 ResNet

The original ResNet implementation consists of two types of blocks, basic and bottleneck, and uses residual connections to avoid the vanishing gradient problem. The first one is used for shallower networks like ResNet-18 and 34, which consist of two sequential 3x3 convolutional layers. The key feature is the identity shortcut, which adds the input to the output of the two convolutional layers. This skip connection allows gradients to flow directly through the network, preventing them from vanishing. The block's production has the same number of channels as the second convolutional layer. For deeper networks, like ResNet-50, the latter block is more efficient. It uses a sequence of three convolutional layers: a 1x1 convolution to reduce dimensionality, a 3x3 convolution, and a final 1x1 convolution to restore dimensionality. This bottleneck design minimises the number of parameters and computation while maintaining representational capacity. The identity shortcut is still present, connecting the block's input to its final output.

Our implementation starts with an initial 3x3 convolution, batch normalisation, an activation function, and a pooling layer. This implementation is designed to handle both RGB and greyscale images by using separate initial convolutional layers. The forward pass intelligently routes the input to the correct layer based on the image's channel count. The feature extraction backbone consists of four main layers, each containing several residual blocks, as defined by the depth parameter. The output channel count of each layer doubles, and the spatial dimensions are halved with a stride of 2 in the first block of layers 2, 3, and 4. A downsampling layer is automatically added to the identity shortcut when the input and output dimensions don't match, ensuring the addition operation is always valid. After the last residual layer, a final pooling operation is applied to produce a fixed-size feature vector.

4.2.4 UNet

Our implementation of UNet is adapted for classification, despite its original purpose being segmentation, by reformulating the symmetrical encoder-decoder architecture and skip connections to create a powerful feature extractor. The UNet architecture has three fundamental blocks: the double convolution, the down and up blocks. The first one is the fundamental convolutional block, consisting of two consecutive 3x3 convolutions, each followed by batch normalisation and a specified activation function. This block is used in both the encoder and decoder parts of the network. The down block consists of a single step in the encoder. It first applies a pooling to downsample the feature map's spatial dimensions, reducing its size by half. A double convolutional block immediately follows this to increase the number of feature channels. As for the up block, it is also a single step in the decoder. It begins with a transposed convolutional layer that upsamples the feature map, doubling its spatial dimensions. A key feature of the U-Net architecture is the subsequent skip connection; the upsampled feature map is concatenated with a feature map of the same size from the corresponding down module in the encoder. This concatenation allows the decoder to leverage fine-grained, high-resolution features from the encoder, which is crucial for preserving spatial details. A double convolutional block is then applied to the concatenated feature maps.

The architecture is designed to handle both RGB and greyscale images. Separate initial double convolutional layers are used to process these different input types. Similarly, the encoder and decoder layers are duplicated for each input type. The forward pass routes each image in a batch to its corresponding encoder-decoder path. The feature-extraction process follows the classic U-Net architecture. The input image is passed through a series of Down modules in the encoder, progressively reducing its spatial size while increasing its channel count. The number of layers in the encoder-decoder blocks is determined using a depth parameter. The resulting low-resolution feature map is then passed through the decoder's up modules, where it is upsampled and fused with high-resolution features from the encoder via skip connections. The decoder's final output is a high-resolution feature map. Unlike a standard U-Net, which would use this output for pixel-wise segmentation, this implementation prepares the decoder's output for classification. It applies an output convolution to reduce the number of channels to a specified size. It then uses adaptive average pooling to downsample this feature map to a fixed size, effectively creating a single, comprehensive feature vector for each image.

4.2.5 ViT

The original ViT architecture is composed of two main components: the patch embedding and the transformer encoder. The first one focuses on splitting an input image into a grid of non-overlapping patches. A convolutional layer projects each patch into a high-dimensional embedding vector. These patch embeddings are then flattened and rearranged into a sequence. A special learnable class token is prepended to this sequence, which serves as the final representation for classification. Finally, positional embeddings are added to the patch embeddings to provide the model with spatial information. This has been adapted as the Transformers are mainly used for NLP tasks. The second main idea consists of two sub-layers, each with a residual connection around it. The first sub-layer is a Multi-Head Self-Attention block, which allows the model to weigh the importance of different patches. The second sub-layer is a position-wise feed-forward network. Each sub-layer is preceded by Layer Normalisation, which helps stabilise training.

Our architecture handles the dynamic input of both RGB and greyscale images. It initialises two separate patch embedding modules. In the forward pass, it selects the correct embedding module based on the channel dimension of each image in the batch. The backbone uses the processed patch embeddings, which are passed through a sequential stack of transformer encoder blocks. The output of this entire encoder stack is a sequence of feature vectors for each patch. For the final classification, only the class token embedding is used, as it has been trained to aggregate information from all the other patches. This class token embedding is normalised and then passed to the classification head.

4.2.6 Swin

The core of Swin lies in its hierarchical structure and Shifted Window Attention, which addresses the computational cost of global self-attention while still enabling information flow across different image regions. Therefore, it is based on several components, including the patch embedding, the window attention, the Swin transformer block, the patch merging and the basic layer. The patch embedding is the same as the ViT model, but a key difference is that Swin doesn't use a class token or absolute positional embeddings by default (though the option for ape is included). Instead, it relies on the hierarchical structure and relative position bias to encode spatial information. Now, the heart of the Swin, that is to say, the window attention mechanism. It performs multi-head self-attention, but unlike a standard Transformer, it restricts the attention calculation to non-overlapping local windows.

This significantly reduces the computational complexity. The module also incorporates a relative-position bias to encode the spatial relationships between tokens within a window. The upcoming fundamental block is the Swin transformer, which consists of two consecutive attention blocks. The first block uses regular window attention, and the second uses Shifted Window Attention. By shifting the windows in the second block, the model allows information to be exchanged between adjacent windows that were previously separate. This shifting mechanism is crucial for enabling cross-window information propagation. Another essential block would be the patch merging. This module downsamples the feature map by merging adjacent patches. It takes four neighbouring patches, concatenates their feature vectors, and applies a linear layer to reduce the feature dimension. This operation effectively halves the spatial resolution while doubling the number of channels, creating a pyramid of feature maps similar to a traditional convolutional neural network. Now, for the last block, called the basic layer, this one represents a single stage in the Swin Transformer's hierarchical structure. It consists of a series of Swin transformer blocks, followed by a patch-merging layer for all but the last stage. The blocks alternate between regular and shifted window attention, ensuring the model learns both local and global features.

Now our implementation respects the fundamentals of Swin but also has some tailor-made properties, such as the ability to handle both RGB and greyscale images. It uses separate patch-embedding modules for each input type to process the initial image patches correctly. At the end of the backbone, the final output is a feature map with a high number of channels and a small spatial size. This implementation uses a pooling function to flatten this feature map into a single feature vector, which serves as the final image-based representation for classification.

4.2.7 Pre-trained Models

The pre-trained models approach uses a technique called TL, where we extract the backbone of a model, i.e., delete unnecessary layers such as the classification head, load the pre-trained weights, freeze these layers, and extract features from the images. Then a specialised block is used to train the weights on our specific task and dataset, instead of relying solely on pre-trained weights. The pre-trained models to be chosen depend on the task at hand; in our case, we need models trained on remote sensing data, specifically satellite imagery. This allows the model to converge much faster, therefore requiring fewer epochs in total to train and validate. In addition, by freezing the backbone, we are effectively telling the optimisation algorithm not to update these weights during training. The model's feature extraction capabilities, the ability to identify low-level and high-level features in an image, are kept constant. Another advantage is improved performance with a small dataset, because training a DL on a small dataset can lead to overfitting, where the model learns the noise in the training data rather than the underlying patterns. The pre-trained backbone already possesses a robust understanding of general image features, which helps the model generalise better to the new task, even with limited training examples. Finally, the introduction of specialised layers that are not frozen allows the model to be trained, enabling it to learn to map these robust features to the desired output classes.

The pre-trained models used for this are the same as the ones used in the previous subsections. The models have slight variations but overall are very similar to the ones described above. The modules where the models come from are either torchgeo [132] for ResNet, UNet, ViT and Swin or Big Earth Net [21, 51] for DenseNet.

4.3 Evaluation Metrics

In DL models, several evaluation metrics are needed to assess their performance. These allow direct comparison between different architectures and facilitate choosing the best one.

The first evaluation metric is the Cross-Entropy Loss, which measures how well the predicted probability distribution matches the true class labels. It quantifies the dissimilarity between the predicted class probabilities and the actual class distribution, penalising confident but incorrect predictions more heavily. This metric is particularly suitable for classification tasks where the model outputs a probability distribution over multiple classes.

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c}) \tag{4.1}$$

where, N: is the number of samples, C is the number of classes, $y_{i,c}$ verifies if i belongs to class c giving it 1 otherwise it is 0 and $\hat{y}_{i,c}$ is the predicted probability for class c. The loss function is also calculated on the validation data, which is unseen.

To comprehensively evaluate the performance of the trained models, multiple metrics were used. These metrics capture different aspects of classification performance and are particularly important when class imbalance or varying misclassification costs exist.

Accuracy measures the overall proportion of correctly classified instances out of all instances. It is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.2}$$

where TP denotes *True Positives*, TN *True Negatives*, FP *False Positives*, and FN *False Negatives*. While accuracy provides a general overview of model performance, it can be misleading in highly imbalanced datasets, where the majority class may dominate it.

Another important metric is Precision, which quantifies the proportion of positive predictions that are actually correct, reflecting the reliability of the model when it predicts a positive outcome:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.3}$$

High precision indicates that the model makes few false positives. This metric is vital in applications where the cost of a false positive is high.

A further criterion is called Recall, which is also referred to as *sensitivity* or the *true positive rate*, and measures the proportion of actual positive instances that are correctly identified:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.4}$$

A high recall means the model captures most positive instances, which is crucial in scenarios where missing a positive case has significant consequences.

As the final metric, the F1-score is the harmonic mean of precision and recall:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.5}$$

It provides a balanced measure that accounts for both false positives and false negatives. The F1-score is especially useful when the dataset is imbalanced, as it does not favour either precision or recall excessively. In this study, the F1-score was selected as the primary criterion for model ranking due to the relatively small dataset and the need to balance precision and recall.

Finally, training time was recorded for each fold, providing insight into computational efficiency. While it is not a direct measure of model quality, it is a relevant practical consideration for model selection when resources are limited.

4.4 Computational Setup

All experiments in this work were conducted on a high-performance workstation equipped with an Intel Core i9-10900K CPU running at 3.70 GHz, 64 GB of DDR4 RAM, an NVIDIA Quadro RTX 8000 GPU supporting 48 GB of GDDR6, and a total disk capacity of 13 TB. This hardware configuration provided sufficient computational power and storage to train and evaluate deep learning models and preprocess datasets.

The software environment was based on Python version 3.11.13 and included several widely used libraries and frameworks. PyTorch version 2.4.1 served as the primary deep learning framework, together with CUDA support version 11.8, complemented by Torchvision version 0.19.1 for computer vision tasks and TorchGeo version 0.7.0 for remote sensing pre-trained models. Data handling and preprocessing were facilitated using NumPy version 1.26.4 and Pandas version 2.2.3, while model evaluation and machine learning utilities were supported by Scikit-learn version 1.6.1. Visualisation of experimental results was carried out using Matplotlib version 3.9.2, and image manipulation was supported by Pillow version 10.4.0. Finally, the models also used a fixed seed for reproducibility, set to 55. This software setup ensured efficient implementation of models and reliable reproducibility of results.

4.5 Summary

Chapter 4 presented the experimental setup, detailing the implementation and evaluation of various DL architectures. It began by outlining the experimental design, including dataset configurations, training procedures, and different data perspectives, as well as the entire lake and in-situ area crops. Both augmented and non-augmented datasets were used to evaluate the effect of data diversity on model performance. Grid Search was used to tune hyperparameters across architectures and systematically identify the best configurations. Learning curves were analysed to monitor overfitting, determine the optimal number of epochs, and ensure training stability.

The chapter also described the DL model architectures, including DenseNet, ResNet, UNet, ViT, and Swin, as well as the use of pre-trained models for TL. Evaluation metrics such as accuracy, precision, recall, and F1-score were used to comprehensively assess performance, with F1-score serving as the primary ranking criterion due to dataset imbalance. Finally, the computational setup was detailed, including the hardware and software environment. Together, these elements provided a robust experimental foundation for comparing architectures and identifying the most effective models for water quality classification.

Chapter 5

Results & Discussion

In this chapter, we will review the results for both lakes and discuss them. Specifically, in sections 5.1 and 5.2, the results for Lake Azul and Lake Fogo are presented, respectively. Ultimately, section 5.3 performs an overall discussion of the obtained results.

5.1 Lake Azul

An overview of the model tuning outcomes derived from the small patch centred on the in-situ collection site of Lake Azul is presented in Table 5.1.

In the first category, we can see that, based on the F1-score, all models performed exceptionally well, with scores ranging from 0.924 for ResNet18 to 0.967 for Swin-Tiny. A notable pattern is the use of a consistent learning rate (q) of 0.001 across all models. While most CNNs use the ReLU activation function (h), the transformer-based models exclusively utilize Gaussian Error Linear Unit (GELU), and interestingly DenseNet uses a softmax activation. The transformer models, specifically ViT and Swin, also introduce unique hyperparameters like embed dim (l), num heads (n), window size (o), and MLP ratio (p), which do not apply to the other models. Additionally, the batch size (a) shows a clear preference for smaller values, with a batch size of 2 used in 5 out of 8 models. The growth rate (b) is most commonly 24, seen in both DenseNet121 and DenseNet40. The dropout rate (f), when specified, is predominantly 0.0, with notable exceptions of the Swin-Tiny and ViT models, which have dropout rates of 0.5 and 0.5 respectively. The transformers also stand out for their unique hyperparameter choices, with both Swin models having an embed dim (l) of 64, a patch size (k) of 6, and an MLP ratio (p) of 4.0. The learning rate (q) and weight decay (r) are remarkably consistent across all models at 0.001 and 0.0001, respectively. For the models that use the pooling function, the average pooling is used, except for DenseNet121, which uses max pooling.

In the subsequent category, the batch sizes (a) of 2 and 4 are used almost equally, with four models

Table 5.1: Best candidate deep learning model results with different configurations based on Small patch of Lake Azul. Legend: **a.** Batch Size, **b.** Growth Rate, **c.** BN Size, **d.** Base Channels, **e.** Bias, **f.** Dropout Rate, **g.** Attention Dropout, **h.** Activation Function, **i.** Pooling, **j.** Batch Norm, **k.** Patch Size, **l.** Embed Dim, **m.** Num Heads, **n.** Window Size, **o.** MLP Ratio, **p.** Learning Rate, **q.** Weight Decay, **r.** λ_{l1}, **s.** F1 Score, **t.** Accuracy, **u.** Recall, **v.** Precision, **w.** Train Time (s)

Model	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
Dataset without Augmentation																							
DenseNet40	4	32	3	–	False	–	–	softmax	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.941	0.946	0.946	0.938	119.639
DenseNet121	2	24	3	–	False	–	–	relu	max	False	–	–	–	–	–	0.001	0.0001	0.0001	0.961	0.961	0.961	0.964	313.200
ResNet18	2	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.924	0.929	0.929	0.932	105.918
ResNet50	2	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.946	0.946	0.946	0.948	167.488
UNet	4	–	–	32	–	0.0	–	relu	–	False	–	–	–	–	–	0.001	0.0001	0.0001	0.946	0.946	0.946	0.950	96.449
Swin-Tiny	2	–	–	–	–	0.5	0.0	gelu	–	–	6	64	–	4	4.0	0.001	0.0001	0.0001	0.964	0.964	0.964	0.964	344.385
Swin-Small	2	–	–	–	–	0.0	0.0	relu	–	–	6	64	–	4	4.0	0.001	0.0001	0.0001	0.967	0.967	0.967	0.970	375.436
ViT	2	–	–	–	–	0.5	–	relu	–	–	5	32	4	–	4.0	0.001	0.0001	0.0001	0.946	0.946	0.946	0.952	201.303
Dataset with Augmentation																							
DenseNet40	2	24	4	–	False	–	–	softmax	max	False	–	–	–	–	–	0.001	0.0001	0.0001	0.979	0.979	0.979	0.98	418.249
DenseNet121	2	32	3	–	False	–	–	softmax	max	False	–	–	–	–	–	0.001	0.0001	0.0001	0.979	0.979	0.979	0.980	590.291
ResNet18	2	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.979	0.979	0.979	0.980	505.05
ResNet50	4	–	–	–	False	0.5	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.936	0.944	0.944	0.933	399.595
UNet	4	–	–	32	–	0.0	–	sigmoid	–	False	–	–	–	–	–	0.001	0.0001	0.0001	0.975	0.978	0.978	0.973	505.302
Swin-Tiny	2	–	–	–	–	0.0	0.0	relu	–	–	6	64	–	4	2.0	0.001	0.0001	0.0001	0.979	0.979	0.979	0.980	623.658
Swin-Small	4	–	–	–	–	0.5	0.1	gelu	–	–	6	64	–	4	2.0	0.001	0.0001	0.0001	0.979	0.979	0.979	0.980	551.633
ViT	4	–	–	–	–	0.0	–	gelu	–	–	5	64	4	–	2.0	0.001	0.0001	0.0001	0.979	0.979	0.979	0.980	248.142
Dataset using Pre-trained models																							
DenseNet	8	–	–	–	–	0.0	–	relu	–	False	–	–	–	–	–	0.001	0.0001	0.0001	0.905	0.911	0.911	0.913	81.355
ResNet18	2	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.910	0.911	0.911	0.922	135.021
ResNet50	2	–	–	–	False	0.5	–	relu	avg	False	–	–	–	–	–	0.0001	0.0001	0.0001	0.689	0.714	0.714	0.704	192.144
UNet	4	–	–	–	–	0.0	–	relu	–	–	–	–	–	–	–	0.001	0.0001	0.0001	0.955	0.955	0.955	0.958	288.316
Swin-Tiny	2	–	–	–	–	0.0	–	relu	–	–	–	–	–	–	–	0.001	0.0001	0.0001	0.979	0.979	0.979	0.980	345.074
ViT	2	–	–	–	–	0.0	–	gelu	–	–	–	–	–	–	–	0.001	0.0001	0.0001	0.970	0.970	0.970	0.971	258.054

using a batch size of 2 and the other four using a size of 4. The growth rate (b) varies between 24 and 32 for the DenseNet models. The dropout rate (f) is again 0.0, except for ResNet50 and Swin-Small, which use 0.5. The transformer models' configurations differ slightly; for example, the Swin-Tiny and Swin-Small models both have a lower MLP ratio (p) of 2.0 compared to the non-augmented dataset. Another curious fact is that the UNet uses a sigmoid activation function, and both DenseNet models use a sigmoid activation function, while the others use either ReLU or GELU. The learning rate (q) and weight decay (r) remain constant at 0.001 and 0.0001, respectively. As for the pooling function, both DenseNet models use max pooling, whereas the ResNet models use avg pooling.

The "Dataset using Pre-trained models"category is characterised by a mix of results and a few key differences in hyperparameters. Most models still use a learning rate (q) of 0.001, but the pre-trained ResNet50 model uses a much smaller value of 0.0001, which could contribute to its poor performance. Similar to the other categories, ReLU is the dominant activation function, but the pre-trained ViT uses GELU. It shows a similar variation in batch size (a), with values of 2, 4, and especially 8 being used for the DenseNet model. The dropout rate (f) is consistently 0.0 for all models, except for ResNet50, which uses 0.5. This category also has a notable outlier in its learning rate (q), ResNet50 is the only model to use a rate of 0.0001, which is an order of magnitude smaller than the 0.001 used by all other models in the Table. The pooling function (i) chosen by the specific models that use them was avg.

When focusing on the F1-score (t), the 'Dataset with Augmentation' category stands out, with six out of eight models achieving a score of 0.979, a substantial improvement over the best scores in the non-augmented category. For example, the chosen comparison metric for DenseNet40 increased from 0.941 to 0.979 with the addition of data augmentation. Similarly, ResNet18's F1-score rose from 0.924 to 0.979.

This highlights the significant benefits of data augmentation for these models.

The 'Dataset using Pre-trained models' category presents a mix of results. While the Swin-Tiny model achieved a high F1-score of 0.979, matching the best-augmented models, the ResNet50 model was poor, achieving 0.689. This is a drastic decrease compared to the 0.946 it achieved without augmentation and the 0.936 it achieved with augmentation, underscoring that pre-training can be a hit-or-miss strategy depending on the model and dataset compatibility.

Regarding the Train Time, it also reveals interesting patterns. As expected, models with augmentation take longer to train. The average training time for models without augmentation is approximately 215 seconds, while the average with augmentation is 480 seconds, more than 2 times as long. Conversely, the pre-trained models have a similar average training time, at approximately 216 minutes with the first technique. While Swin-Tiny with pre-training achieves an F1-score of 0.979 with a train time of 345 seconds, the same F1-score is achieved with augmentation in 623 seconds. This shows that for the same top-tier performance, using a pre-trained Swin-Tiny model requires approximately 45% less training time.

The following Figure 5.1 shows the F1-score performance of models trained on the small dataset of Lake Azul without any data augmentation. The results highlight the limitations of working with reduced data availability, where models tend to underperform due to the lack of variability in the training set.

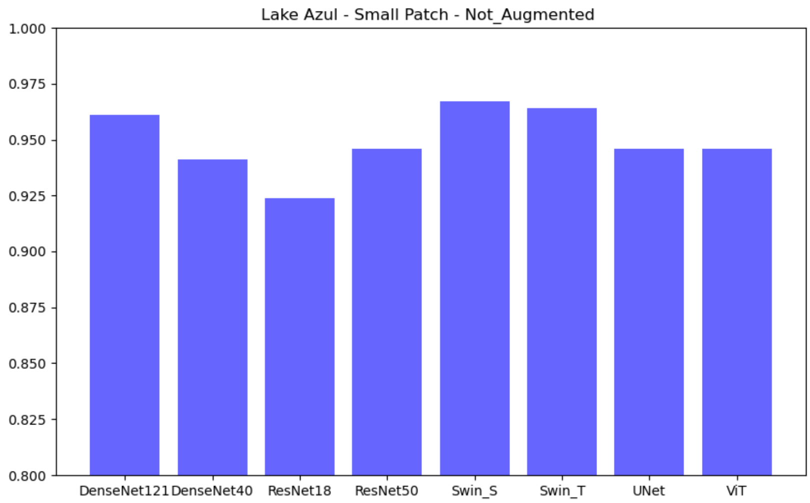

Figure 5.1: Small Not Augmented Lake Azul

Here, in Figure 5.2, the use of data augmentation on the small Lake Azul dataset clearly improves F1 Scores. The synthetic variability introduced during training helps models generalise better, indicating that augmentation is particularly effective when only limited samples are available.

This Figure 5.3 illustrates the performance of TL on the small Lake Azul dataset. Leveraging pre-trained weights yields a noticeable boost in F1-scores compared to training from scratch, confirming that TL is highly beneficial when data is scarce.

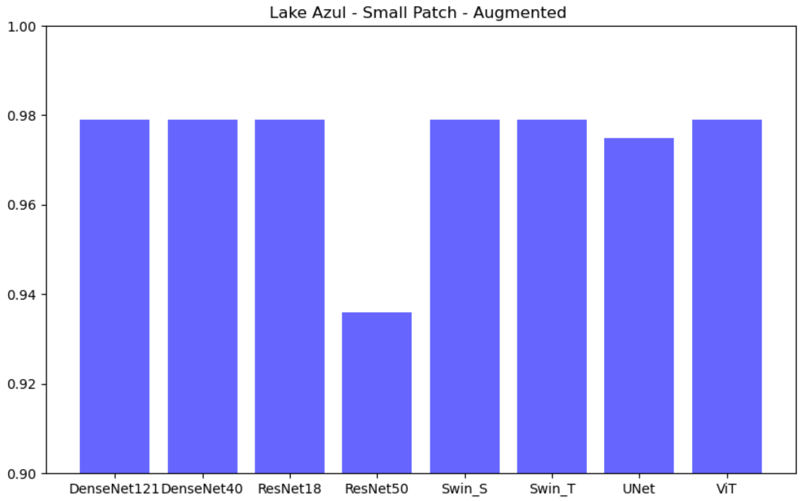

Figure 5.2: Small Augmented Lake Azul

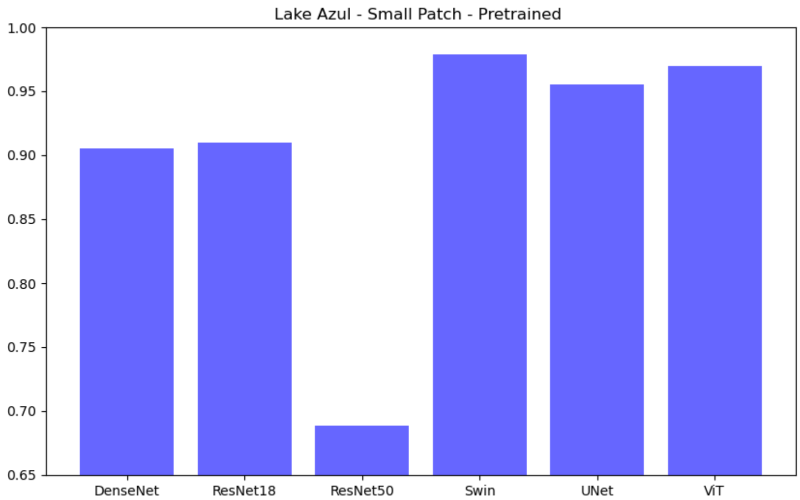

Figure 5.3: Small Pretrained Lake Azul

Overall, for Lake Azul with the Small Patch dataset, the results show that models trained without augmentation performed the weakest, confirming the limitations of scarce data. Augmentation provided substantial improvement, while TL achieved the highest F1 Scores, demonstrating that pretrained models can adapt effectively even under data scarcity.

The tuning performance obtained using the full satellite image covering the entire Lake Azul is summarised in Table 5.2.

In the 'Dataset without Augmentation' category, there's a mix of batch sizes (a), with 2 and 4 perfectly split. The growth rate (b) varies, with DenseNet40 at 24 and DenseNet121 at 32. The dropout rate (f), when used, is 0.0 for ResNet18, UNet and Swin-Tiny, and 0.5 for ResNet50 and ViT. The pooling (i) functions used

Table 5.2: Best candidate deep learning model results with different configurations based on the Full image of Lake Azul. Legend: **a.** Batch Size, **b.** Growth Rate, **c.** BN Size, **d.** Base Channels, **e.** Bias, **f.** Dropout Rate, **g.** Attention Dropout, **h.** Activation Function, **i.** Pooling, **j.** Batch Norm, **k.** Patch Size, **l.** Embed Dim, **m.** Num Heads, **n.** Window Size, **o.** MLP Ratio, **p.** Learning Rate, **q.** Weight Decay, **r.** λ_{l1}, **s.** F1 Score, **t.** Accuracy, **u.** Recall, **v.** Precision, **w.** Train Time (s)

Model	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
Dataset without Augmentation																							
DenseNet40	4	24	3	–	False	–	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.958	0.958	0.958	0.959	112.969
DenseNet121	4	32	3	–	False	–	–	relu	max	False	–	–	–	–	–	0.001	0.0001	0.0001	0.952	0.952	0.952	0.959	433.408
ResNet18	2	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.943	0.952	0.952	0.935	290.733
ResNet50	4	–	–	–	False	0.5	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.934	0.938	0.938	0.937	212.168
UNet	2	–	–	64	–	0.0	–	relu	–	False	–	–	–	–	–	0.001	0.0001	0.0001	0.962	0.964	0.964	0.959	285.143
Swin-Tiny	2	–	–	–	–	0.0	0.0	relu	–	–	16	96	–	8	4.0	0.001	0.0001	0.0001	0.970	0.970	0.970	0.970	187.261
Swin-Small	2	–	–	–	–	0.5	0.0	gelu	–	–	16	64	–	8	2.0	0.001	0.0001	0.0001	0.967	0.967	0.967	0.970	453.621
ViT	4	–	–	–	–	0.5	–	relu	–	–	16	32	–	4	4.0	0.001	0.0001	0.0001	0.949	0.949	0.949	0.953	180.435
Dataset with Augmentation																							
DenseNet40	4	32	4	–	False	–	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.979	0.979	0.979	0.980	207.067
DenseNet121	2	32	3	–	False	–	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.971	0.971	0.971	0.972	461.173
ResNet18	2	–	–	–	False	0.0	–	softmax	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.965	0.965	0.965	0.966	526.882
ResNet50	2	–	–	–	False	0.0	–	softmax	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.958	0.958	0.958	0.962	995.785
UNet	2	–	–	64	–	0.0	–	relu	–	False	–	–	–	–	–	0.001	0.0001	0.0001	0.944	0.948	0.948	0.947	429.425
Swin-Tiny	2	–	–	–	–	0.0	0.1	gelu	–	–	16	64	–	8	4.0	0.001	0.0001	0.0001	0.968	0.968	0.968	0.971	401.468
Swin-Small	2	–	–	–	–	0.5	0.0	relu	–	–	16	96	–	8	4.0	0.001	0.0001	0.0001	0.979	0.979	0.979	0.980	369.03
ViT	2	–	–	–	–	0.0	–	gelu	–	–	16	32	8	–	4.0	0.001	0.0001	0.0001	0.968	0.968	0.968	0.969	223.861
Dataset using Pre-trained models																							
DenseNet	8	–	–	–	–	0.0	–	relu	–	False	–	–	–	–	–	0.001	0.0001	0.0001	0.967	0.967	0.967	0.970	86.886
ResNet18	8	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.933	0.940	0.940	0.928	279.335
ResNet50	8	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.686	0.619	0.619	0.792	177.214
UNet	8	–	–	–	–	0.0	–	relu	–	–	–	–	–	–	–	0.001	0.0001	0.0001	0.961	0.961	0.961	0.967	385.158
Swin-Tiny	2	–	–	–	–	0.0	–	gelu	–	–	–	–	–	–	–	0.0001	0.0001	0.0001	0.956	0.958	0.958	0.954	350.387
ViT	4	–	–	–	–	0.0	–	gelu	–	–	–	–	–	–	–	0.001	0.0001	0.0001	0.955	0.955	0.955	0.955	120.935

across the models are mainly avg, but DenseNet121 uses max. The activation (h) functions are all ReLU, but with a clear exception being Swin-Tiny. The transformer models again use unique hyperparameters, such as a patch size (k) of 16 and a num heads (n) of 8. The learning rate (q) and weight decay (r) are consistently 0.001 and 0.0001, respectively, across all models.

In the 'Dataset with Augmentation' category, the batch size (a) is predominantly 2, used by seven out of eight models, where the DenseNet40 is the clear exception. The growth rate (b) is 32 for both DenseNet models, but the DenseNet40 uses a BN size of 4, while the other uses 3. The dropout rate (f) is 0.0 for most models, except for Swin-Small, which uses 0.5. Swin-Tiny uses an attention dropout (g) of 0.1. The DenseNet, UNet, and Swin-Small all use the ReLU activation function; meanwhile, the ResNet models use softmax and the Swin-Tiny and ViT use the GELU activation function. All the models that use pooling use average pooling. The transformer models continue to use a patch size (k) of 16. The ViT model has 8 num_heads. The learning rate (q) and weight decay (r) remain at 0.001 and 0.0001, respectively.

For the 'Dataset using Pre-trained models' category, the batch size (a) is notably larger, with 8 being the most common value, used by four models. The ResNet models both use the avg pooling (i) function. The dropout rate (f) is consistently 0.0 for all models where it is specified. The learning rate (q) is predominantly 0.001, but Swin-Tiny uses a smaller rate of 0.0001, which is an order of magnitude less. All pre-trained models have no specified patch size or other transformer-specific parameters, as the pre-trained architecture determines these.

When evaluating performance based on the chosen metric, the 'Dataset with Augmentation' category shows the best results again. The DenseNet40 and Swin-Small models both achieved the highest F1-score of 0.979. This is a significant improvement over their non-augmented counterparts, which had for the same metric a result of 0.958 and 0.967, respectively. When looking at the other models and comparing their performance, the same happens again: for example, the ResNet models show an increase in performance; the same can be said for UNet and ViT. This clearly demonstrates that data augmentation is a very effective strategy for boosting model performance on this dataset.

The 'Dataset using Pre-trained models' category yields mixed results again. The ResNet50 model performs poorly with a very low F1-score of 0.686. This is a dramatic drop compared to the 0.934 F1-score it achieved without augmentation. On the other hand, DenseNet and UNet perform well, with F1-score of 0.967 and 0.961, respectively, which are very competitive with the top-performing augmented models.

Comparing the train time, models with augmentation take significantly longer to train. The average train time for the non-augmented dataset is approximately 269 seconds, whereas the average for the augmented dataset is about 452 seconds, a 68% increase. The pre-trained models have a much shorter average train time of approximately 233 seconds, which is a key advantage. For instance, the Swin-Small model with augmentation achieves an F1-score of 0.979 with a train time of 369 seconds. However, the Swin-Tiny with pre-training achieves a slightly lower but still competitive F1-score of 0.956 in 350 seconds, demonstrating a balance between performance and training efficiency.

With the larger Lake Azul dataset, models trained without augmentation already achieve more stable F1 scores than in the small-dataset scenario, as shown in Figure 5.4. However, the absence of augmentation still limits peak performance and leaves models more prone to overfitting.

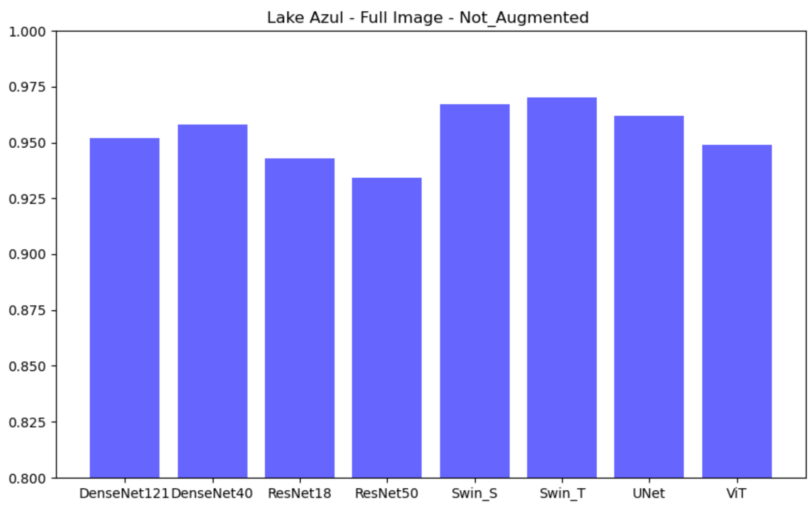

Figure 5.4: Big Not Augmented Lake Azul

The following Figure 5.5 shows that applying augmentation on the big dataset further increases F1-scores. Even with more data available, augmentation continues to provide gains by diversifying training samples and enhancing generalisation.

The pre-trained models applied to the large Lake Azul dataset achieve the highest F1 Scores among the three approaches, as shown in Figure 5.6. This confirms that combining larger datasets with TL yields

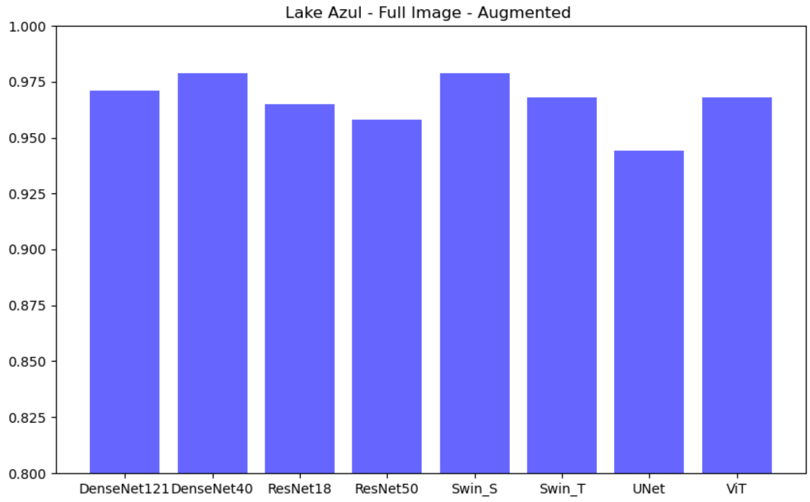

Figure 5.5: Big Augmented Lake Azul

the most effective framework for water quality classification.

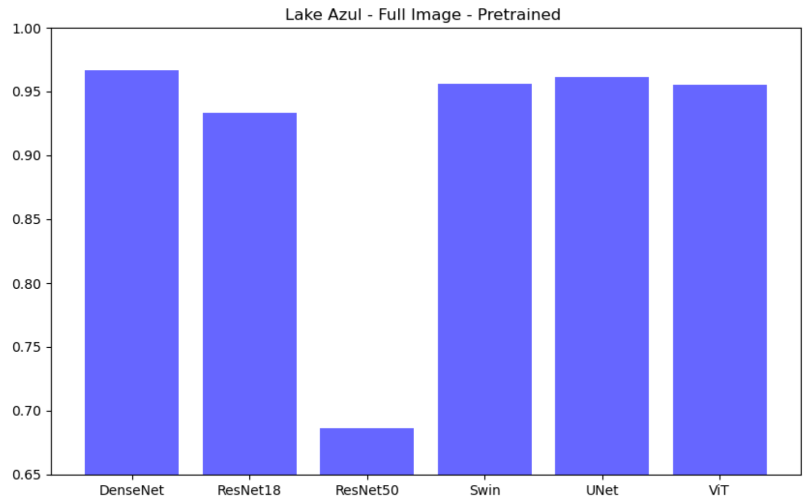

Figure 5.6: Big Pretrained Lake Azul

When scaling to the Lake Azul Full Image dataset, performance improved across all approaches due to greater data availability. Nevertheless, augmentation still enhanced robustness, and TL once again produced the strongest results, confirming that pretrained models scale efficiently with larger datasets.

5.2 Lake Fogo

The results corresponding to the model tuning performed on the small image patch around the in-situ sampling area of Lake Fogo are shown in Table 5.3.

Table 5.3: Best candidate deep learning model results with different configurations based on the Small Patch of Lake Fogo. Legend: **a.** Batch Size, **b.** Growth Rate, **c.** BN Size, **d.** Base Channels, **e.** Bias, **f.** Dropout Rate, **g.** Attention Dropout, **h.** Activation Function, **i.** Pooling, **j.** Batch Norm, **k.** Patch Size, **l.** Embed Dim, **m.** Num Heads, **n.** Window Size, **o.** MLP Ratio, **p.** Learning Rate, **q.** Weight Decay, **r.** λ_{l1}, **s.** F1 Score, **t.** Accuracy, **u.** Recall, **v.** Precision, **w.** Train Time (s)

Model	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
Dataset without Augmentation																							
DenseNet40	2	32	3	–	False	–	–	softmax	max	False	–	–	–	–	–	0.001	0.0001	0.0001	0.837	0.848	0.848	0.860	129.755
DenseNet121	2	24	3	–	False	–	–	relu	max	False	–	–	–	–	–	0.001	0.0001	0.0001	0.834	0.848	0.848	0.843	195.715
ResNet18	2	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.870	0.878	0.878	0.879	126.781
ResNet50	2	–	–	–	False	0.0	–	softmax	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.853	0.863	0.863	0.863	464.898
UNet	2	–	–	64	–	0.0	–	relu	–	False	–	–	–	–	–	0.001	0.0001	0.0001	0.936	0.938	0.938	0.938	56.964
Swin-Tiny	2	–	–	–	–	0.0	0.0	gelu	–	–	6	96	–	4	4.0	0.001	0.0001	0.0001	0.905	0.908	0.908	0.928	164.951
Swin-Small	2	–	–	–	–	0.5	0.0	relu	–	–	6	64	–	4	2.0	0.001	0.0001	0.0001	0.900	0.908	0.908	0.909	228.445
ViT	2	–	–	–	–	0.5	–	gelu	–	–	5	32	8	-	2.0	0.001	0.0001	0.0001	0.841	0.854	0.854	0.859	71.535
Dataset with Augmentation																							
DenseNet40	2	24	4	–	False	–	–	softmax	max	False	–	–	–	–	–	0.001	0.0001	0.0001	0.962	0.965	0.965	0.962	363.090
DenseNet121	2	32	3	–	False	–	–	softmax	max	False	–	–	–	–	–	0.001	0.0001	0.0001	0.952	0.952	0.952	0.957	617.996
ResNet18	2	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.950	0.957	0.957	0.955	466.123
ResNet50	2	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.926	0.930	0.930	0.930	484.682
UNet	2	–	–	32	–	0.0	–	relu	–	False	–	–	–	–	–	0.001	0.0001	0.0001	0.948	0.948	0.948	0.958	81.410
Swin-Tiny	2	–	–	–	–	0.0	0.1	relu	–	–	6	64	–	4	4.0	0.001	0.0001	0.0001	0.950	0.951	0.951	0.957	285.851
Swin-Small	2	–	–	–	–	0.5	0.0	relu	–	–	6	96	–	4	2.0	0.001	0.0001	0.0001	0.947	0.948	0.948	0.954	394.235
ViT	2	–	–	–	–	0.0	–	relu	–	–	5	32	8	–	2.0	0.001	0.0001	0.0001	0.980	0.980	0.980	0.982	224.424
Dataset using Pre-trained models																							
DenseNet	4	–	–	–	–	0.0	–	relu	–	False	–	–	–	–	–	0.001	0.0001	0.0001	0.935	0.935	0.935	0.944	92.122
ResNet18	4	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.917	0.920	0.920	0.932	67.084
ResNet50	2	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.847	0.860	0.860	0.869	56.805
UNet	4	–	–	–	–	0.0	–	relu	–	–	–	–	–	–	–	0.001	0.0001	0.0001	0.984	0.985	0.985	0.985	1028.080
Swin-Tiny	4	–	–	–	–	0.0	–	relu	–	–	–	–	–	–	–	0.001	0.0001	0.0001	0.946	0.943	0.943	0.957	153.191
ViT	8	–	–	–	–	0.0	–	relu	–	–	–	–	–	–	–	0.001	0.0001	0.0001	0.915	0.920	0.920	0.932	53.909

In the 'Dataset without Augmentation' category, a consistent batch size (a) of 2 is used across all models. The growth rate (b) is set to 32 for DenseNet40 and set to 24 for DenseNet121. For both models, the BN size (c) is 3. The dropout rate (f) is set to 0.0, but the Swin-Small and ViT use 0.5. For the activation function (h), three are used, where DenseNet40 and ResNet50 use the softmax one, the Swin-Tiny and ViT utilise the GELU, while the others use ReLU. Another hyperparameter is the pooling function (i), which either uses max for the DenseNet models and uses avg for the ResNet ones. The learning rate (q) is uniform at 0.001. The patch size (k) is small: ViT uses 5, and Swin models use 6. The Embed Dim (l) is set to 96 for Swin-Tiny, to 64 for Swin-Small and to 32 for ViT.

In the 'Dataset with Augmentation' category, the batch size (a) remains consistently 2 for all models. The growth rate (b) is set to 24 for DenseNet40 and to 32 for DenseNet121, which are the opposite of the previous category. The BN Size is the same as the previous one, but for the DenseNet40 is 4 this time. For the dropout rate (f), all are set to 0.0, except for Swin-Small, which is set to 0.5. For the Swin-Tiny, the attention dropout is set to 0.1, while for Swin-Small it is set to 0.0. All activation (h) functions are set to ReLU; for the DenseNet models, they are set to softmax. As for the pooling (i) functions, these are the same as for the previous category. The learning rate (q) is unchanged from the non-augmented dataset at 0.001.

The 'Dataset using Pre-trained models' category shows a larger variation in batch size (a), with values of 2, 4, and 8. The dropout rate (f) is consistently 0.0 for all models where it is specified. The learning rate (q) is uniform at 0.001. All models in this category use the ReLU activation (h) function. The ResNet models use average pooling (i).

When evaluating performance, the 'Dataset with Augmentation' category generally yields, again, the best results. The ViT model achieved the highest F1-score of 0.980, with a near-perfect performance also reflected in its accuracy, recall, and precision, all at or above 0.980. This is a dramatic increase from its F1-score of 0.841 without augmentation. Other models also see significant improvements, such as DenseNet40, which goes from an F1-score of 0.837 without augmentation to 0.962 with augmentation. The balanced nature of the performance is evident across the board: the F1-score, accuracy, recall, and precision for models in this category are very similar, indicating a strong ability to correctly identify positive cases without making many false positives.

The 'Dataset using Pre-trained models' category presents a mixed bag of results. The UNet model performed exceptionally well, achieving the highest F1-score in the entire Table at 0.984. Its performance is also highly balanced, with accuracies of 0.985, recall of 0.985, and precision of 0.985. This is a substantial improvement over its non-augmented F1-score of 0.936. In contrast, the ResNet50 model, despite using a pre-trained model, performed poorly with an F1-score of 0.847, which is even lower than its non-augmented performance of 0.853. A closer look reveals slightly lower accuracy and recall (0.860) and marginally higher precision (0.869). Other models, such as ResNet18 and ViT, also showed lower performance with pre-training compared to augmentation, with F1-score of 0.917 and 0.915, respectively.

Whereas pre-training can yield exceptional results, as seen with UNet, it is not a guaranteed method for improvement. It can even lead to decreased performance, as evidenced by ResNet50. The performance of the UNet model with pre-training is a strong outlier, requiring a significantly longer train time of 1028.080 seconds compared to the average of 85 seconds for the other pre-trained models. This suggests that achieving a top-tier result with pre-training may entail a high computational cost. As for the different techniques, the first averages the train time at around 180 seconds, while the second averages it at around 304 seconds.

Similar to Figure 5.1, this plot 5.7 shows the baseline performance for Lake Fogo with the small dataset and no augmentation. Again, the models face clear limitations, with relatively low F1 Scores that reflect difficulties in capturing the complexity of water-quality signals from limited data.

Introducing augmentation into the small Lake Fogo dataset substantially improves performance. The F1-scores, in Figure 5.8, demonstrate that even simple augmentation strategies can mitigate overfitting and enhance the robustness of models.

The TL results for the small Lake Fogo dataset emphasise the advantages of leveraging large-scale pre-trained models. The F1-scores obtained in Figure 5.9 generally surpass both non-augmented and augmented configurations, showing that pretrained models adapt effectively to domain-specific tasks.

For Lake Fogo with the Small Patch dataset, a similar trend was observed; without augmentation, models struggled, while augmentation helped compensate for limited variability. TL outperformed both, underlining its importance when data is constrained.

The performance achieved from tuning models on the complete satellite image of Lake Fogo is detailed in Table 5.4.

In the 'Dataset without Augmentation' category, the batch size (a) is consistently 2 across all models

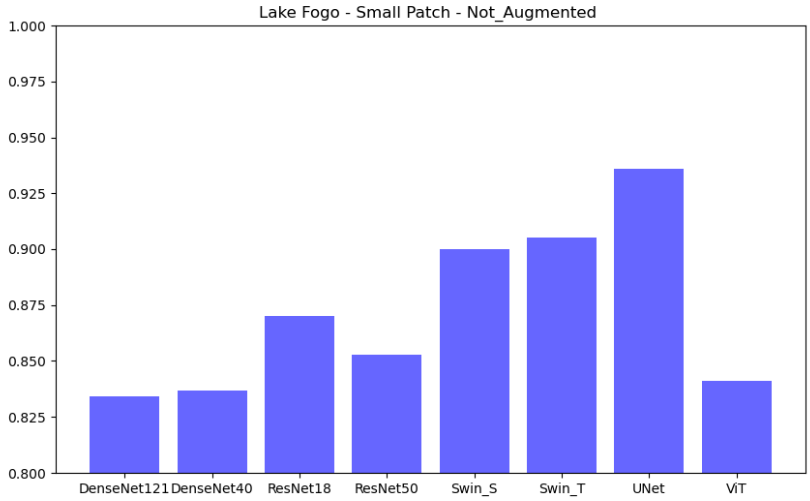

Figure 5.7: Small Not Augmented Lake Fogo

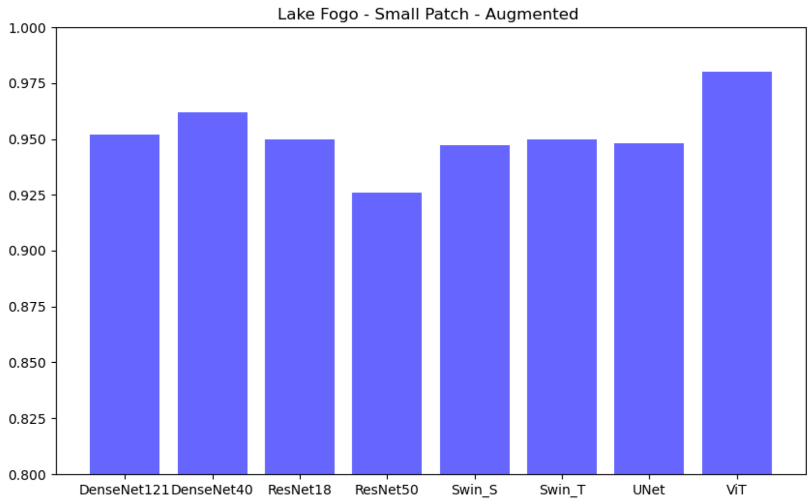

Figure 5.8: Small Augmented Lake Fogo

except for DenseNet121, which uses 4. All models share a standard learning rate (q) of 0.001. The activation (h) function is a mix of ReLU for most of the models, GELU for ViT and softmax for the DenseNet models. For these, they also use the Growth Rate (b) of 24 and for the BN size (c) 4 for the first one and 3 for the second one. The pooling (i) function is also kept at max for these models, while for the ResNet models, it is avg. For dropout (f), ResNet18, ResNet50, UNet and Swin-Tiny use a rate of 0.0, while the transformer models Swin-Small and ViT use 0.5. For the transformer models, they use a patch size (k) of 16. Swin-Tiny uses 96 as the Embed Dim, and the other two use 64. The MLP ratio is kept at 4.0 except for Swin-Small.

For the 'Dataset with Augmentation' category, the batch size (a) is still low, with values of 2 and

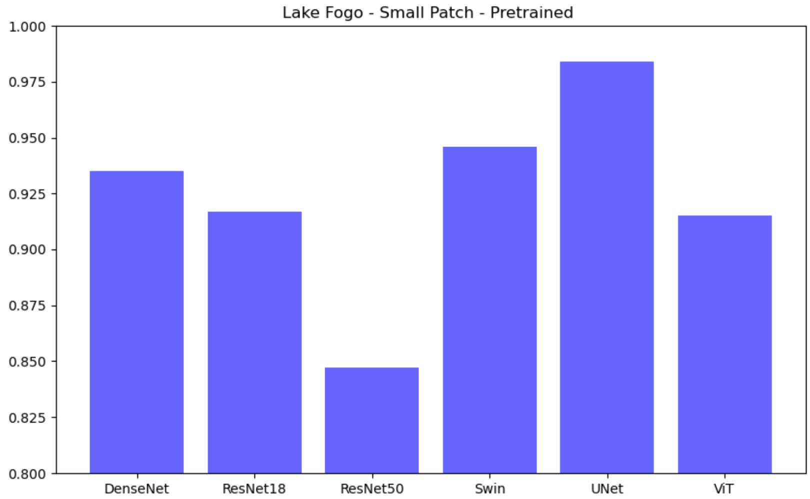

Figure 5.9: Small Pretrained Lake Fogo

Table 5.4: Best candidate deep learning model results with different configurations based on the Full image of Lake Fogo. Legend: **a.** Batch Size, **b.** Growth Rate, **c.** BN Size, **d.** Base Channels, **e.** Bias, **f.** Dropout Rate, **g.** Attention Dropout, **h.** Activation Function, **i.** Pooling, **j.** Batch Norm, **k.** Patch Size, **l.** Embed Dim, **m.** Num Heads, **n.** Window Size, **o.** MLP Ratio, **p.** Learning Rate, **q.** Weight Decay, **r.** λ_{l1}, **s.** F1 Score, **t.** Accuracy, **u.** Recall, **v.** Precision, **w.** Train Time (s)

Model	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**n**	**o**	**p**	**q**	**r**	**s**	**t**	**u**	**v**	**w**
Dataset without Augmentation																							
DenseNet40	2	24	4	–	False	–	–	softmax	max	False	–	–	–	–	–	0.001	0.0001	0.0001	0.865	0.869	0.869	0.894	404.452
DenseNet121	4	24	3	–	False	–	–	softmax	max	False	–	–	–	–	–	0.001	0.0001	0.0001	0.895	0.899	0.899	0.905	1588.497
ResNet18	2	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.852	0.857	0.857	0.887	1114.125
ResNet50	2	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.907	0.908	0.908	0.922	2177.818
UNet	2	–	–	64	–	0.0	–	relu	–	False	–	–	–	–	–	0.001	0.0001	0.0001	0.926	0.926	0.926	0.936	83.635
Swin-Tiny	2	–	–	–	–	0.0	0.0	relu	–	–	16	96	–	8	4.0	0.001	0.0001	0.0001	0.935	0.938	0.938	0.944	1020.308
Swin-Small	2	–	–	–	–	0.5	0.0	relu	–	–	16	64	–	8	2.0	0.001	0.0001	0.0001	0.907	0.911	0.911	0.923	1311.584
ViT	2	–	–	–	–	0.5	–	gelu	–	–	16	64	4	-	4.0	0.001	0.0001	0.0001	0.880	0.890	0.890	0.874	86.246
Dataset with Augmentation																							
DenseNet40	4	32	4	–	False	–	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.966	0.967	0.967	0.970	261.302
DenseNet121	2	24	4	–	False	–	–	softmax	max	False	–	–	–	–	–	0.001	0.0001	0.0001	0.986	0.986	0.986	0.987	4246.412
ResNet18	4	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.978	0.978	0.978	0.979	1647.189
ResNet50	2	–	–	–	False	0.0	–	softmax	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.925	0.927	0.927	0.933	5327.476
UNet	4	–	–	64	–	0.0	–	tanh	–	False	–	–	–	–	–	0.001	0.0001	0.0001	0.950	0.953	0.953	0.955	175.142
Swin-Tiny	2	–	–	–	–	0.5	0.1	gelu	–	–	16	96	–	8	2.0	0.001	0.0001	0.0001	0.984	0.983	0.983	0.987	1304.176
Swin-Small	2	–	–	–	–	0.5	0.0	gelu	–	–	16	64	–	8	4.0	0.001	0.0001	0.0001	0.981	0.982	0.982	0.983	2081.144
ViT	2	–	–	–	–	0.0	–	relu	–	–	8	32	16	–	2.0	0.001	0.0001	0.0001	0.982	0.985	0.985	0.979	916.874
Dataset using Pre-trained models																							
DenseNet	8	–	–	–	–	0.0	–	relu	–	False	–	–	–	–	–	0.001	0.0001	0.0001	0.981	0.982	0.982	0.984	100.252
ResNet18	2	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.938	0.932	0.932	0.956	449.714
ResNet50	8	–	–	–	False	0.0	–	relu	avg	False	–	–	–	–	–	0.001	0.0001	0.0001	0.862	0.866	0.866	0.866	165.881
UNet	8	–	–	–	–	0.0	–	relu	–	–	–	–	–	–	–	0.001	0.0001	0.0001	0.874	0.884	0.884	0.905	219.146
Swin-Tiny	2	–	–	–	–	0.0	–	gelu	–	–	–	–	–	–	–	0.001	0.0001	0.0001	0.957	0.961	0.961	0.958	500.415
ViT	2	–	–	–	–	0.0	–	gelu	–	–	–	–	–	–	–	0.001	0.0001	0.0001	0.947	0.946	0.946	0.953	206.346

4. The Growth Rate (b) increased for DenseNet40 model to 32, while DenseNet121 remains at 24. For both DenseNet models, the BN Size (c) are set to 4. Dropout rates (f) are also varied, with 0.0 and 0.5 being used for the Swin models. Specifically, for the Swin-Tiny, its Attention Dropout (g) is set to 0.1. The activation functions (h) are a mix of ReLU, softmax, and Tangent (Tanh), with UNet using tanh. The deeper variations of the DL models opted for the softmax, while the more shallower ones went for ReLU. The pooling (i) functions are mostly avg, except for DenseNet121, which uses max. For the transformer-specific hyperparameters, these use a Patch Size (k) of 8, while the Embed Dim (l) is set to 96 for Swin-Tiny, 64

for Swin-Small and 32 for ViT. For the MLP ratio (p), it is set to 2.0; only for Swin-Small, it is set to 4.0. The learning rate (q) remains at 0.001.

The 'Dataset using Pre-trained models' category shows a higher batch size (a), with 8 used by DenseNet, ResNet50, and UNet; meanwhile, the others use 2. The dropout rate (f) is consistently 0.0. The activation (h) function is mainly set to ReLU, while for Swin-Tiny and ViT, it is set to GELU. The pooling function is constantly avg. The learning rate (q) and weight decay (r) remain unchanged.

When evaluating performance using F1-score, accuracy, recall, and precision, the 'Dataset with Augmentation' category stands out for its high metrics. DenseNet121 achieved an outstanding F1-score of 0.986, with a similarly high accuracy, recall, and precision all at or above 0.986. This is a significant improvement from its F1-score of 0.895 without augmentation. Similarly, Swin-Tiny achieved an F1-score of 0.984, and Swin-Small a score of 0.981, both with very similar values across all other metrics. This consistency across metrics highlights the models' strong and balanced ability to correctly classify images. This exceptional performance, however, comes at a high computational cost, with a train time of over 4246 seconds. Similarly, Swin-Tiny and Swin-Small achieve excellent metrics with F1-score of 0.984 and 0.981, respectively, but with long training times of 1304 seconds and 2081 seconds. ResNet50 has the longest training time in the entire Table at over 5327 minutes for an F1-score of 0.925.

The 'Dataset without Augmentation' category has generally good performance, but with lower overall metrics compared to the augmented set. The UNet model performed best here, achieving a F1-score of 0.926, an accuracy of 0.926, a recall of 0.926, a precision of 0.936, and a remarkably short training time of only 83 seconds. ViT also shows high efficiency, achieving an F1-score of 0.880 in just 86 seconds. ResNet50 and the Swin models had very long training times, 2177 seconds and 1311 seconds, respectively, for F1-score of 0.907.

The 'Dataset using Pre-trained models' category presents a compelling balance of performance and efficiency. DenseNet achieved a very high and balanced performance, with an F1-score of 0.981, accuracy and recall of 0.982, and a precision of 0.984. This performance is comparable to the best models in the augmented set but with a drastically lower train time of just over 100 seconds, making it an extremely efficient choice. In contrast, ResNet50 and UNet performed poorly with F1-score of 0.862 and 0.874, respectively, despite their relatively short training times of 165 seconds and 219 seconds. This shows that pre-training is not universally effective and can sometimes lead to lower performance than training from scratch, as seen with UNet, which scored 0.926 without pre-training. Overall, while augmentation tends to produce the best results, using a pre-trained model can be a very efficient alternative if the pre-trained weights are a good fit for the dataset.

For Lake Fogo, the larger dataset without augmentation yields stable but moderate F1 Scores, as shown in Figure 5.10. The models capture general trends better than with small datasets, but they do not fully address variability and edge cases.

The impact of augmentation on the large Lake Fogo dataset is evident, with higher F1-scores than in the non-augmented case, demonstrated in Figure 5.11. This demonstrates that augmentation contributes positively regardless of dataset size, although the improvements are less dramatic than with the small

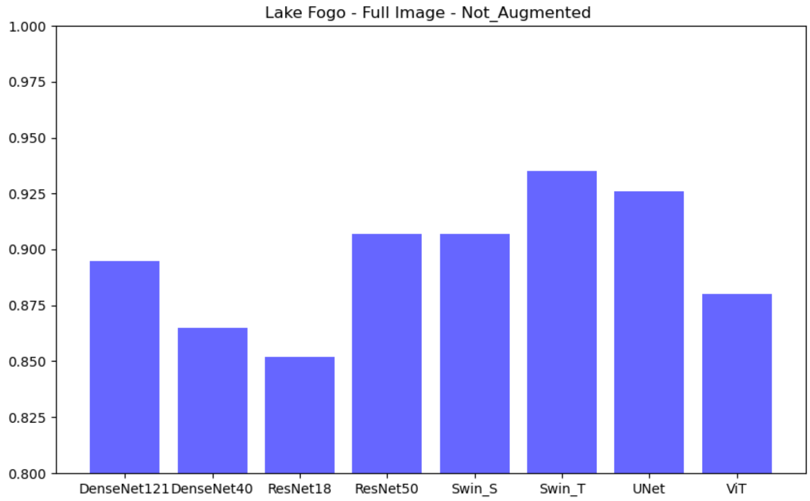

Figure 5.10: Big Not Augmented Lake Fogo

dataset.

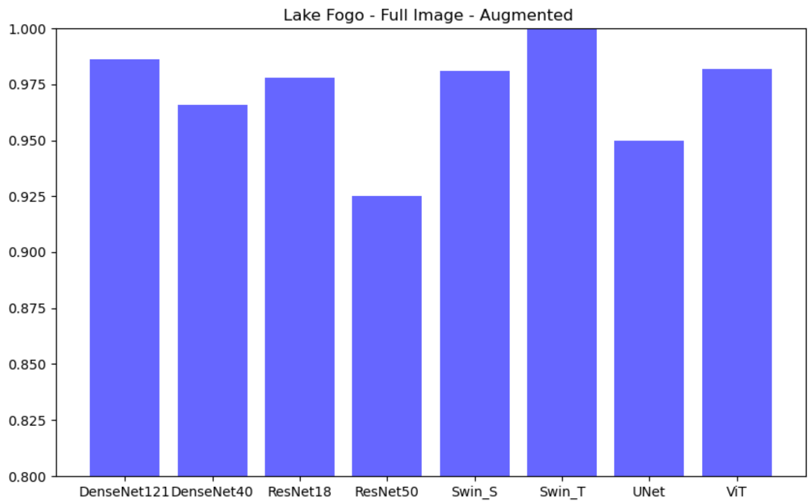

Figure 5.11: Big Augmented Lake Fogo

Finally, the pre-trained models applied to the large Lake Fogo dataset consistently achieve the highest F1-scores, as shown in Figure 5.12. This outcome highlights the harmony between TL and richer datasets, establishing it as the best-performing configuration overall.

Finally, the Lake Fogo big dataset followed the same hierarchy. Models trained without augmentation performed reasonably, augmentation added further improvements, and TL consistently yielded the highest F1-scores. This pattern demonstrates the generalisability of the framework, with consistent benefits across different lakes and dataset sizes.

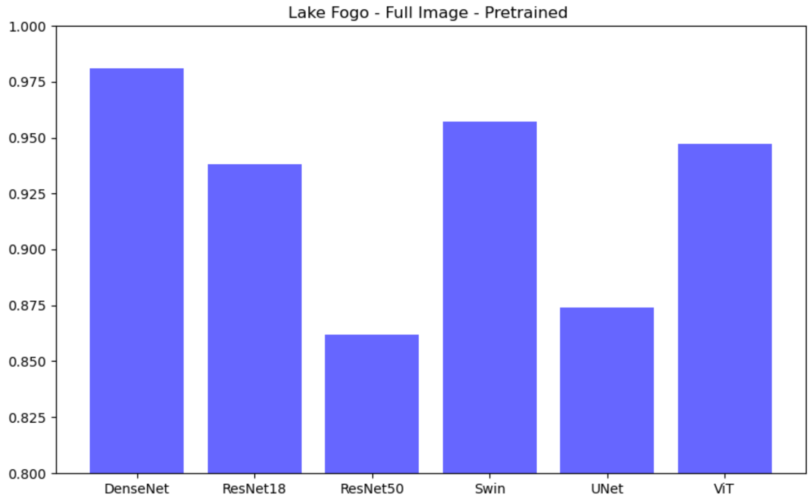

Figure 5.12: Big Pretrained Lake Fogo

5.3 Summary

In summary, the analysis highlights that data augmentation is the most effective strategy for achieving high performance. DL models trained with augmentation consistently outperform their non-augmented counterparts, often reaching near-perfect F1-score and other metrics. This boost in performance, however, comes at the cost of significantly longer training times, in some cases exceeding 1,000 seconds.

Conversely, using pre-trained models presents a high-risk, high-reward scenario. While some pre-trained models like DenseNet achieve top-tier performance comparable to augmented models but with a drastically lower training time, others like ResNet50 perform poorly, even worse than when trained from scratch. This indicates that pre-training effectiveness is highly dependent on the model and the specific dataset.

The analysis also points out that training from scratch without augmentation offers a baseline performance. While the metrics may be lower than those of the other two methods, some models, such as UNet and ViT, demonstrate remarkable efficiency, achieving decent scores with very short training times. This can be a viable option when computational resources or time are limited.

The analysis clearly illustrates a classic trade-off. Data augmentation is the king of performance, consistently delivering the highest F1-score, but at a considerable time cost. The pre-trained approach, while less consistent, offers the best blend of performance and efficiency when it works well, as seen with the DenseNet model's ability to achieve a top-tier F1-score in just 100 seconds. This is a critical insight for real-world applications, where time and computational resources are often limited.

The most intriguing aspect is the erratic performance of pre-trained models. The drastic drop in ResNet50's F1-score from 0.934, without augmentation, to 0.686, with pre-training, is a strong warning that pre-training is not the go-to technique to be used. This phenomenon could be due to a domain shift, where the features learned on the large-scale pre-training dataset are not relevant to the specific characteristics of the current dataset, particularly those of water bodies. The exceptional performance of the pre-trained DenseNet model, on the other hand, suggests a strong feature overlap, making it a suitable candidate for TL.

Additionally, tuning the models is a very effective way to extract the best performance from DL models, where activation functions, pooling methods, and even batch size vary significantly across models and training scenarios, highlighting the need for model-specific tuning rather than a one-size-fits-all approach.

A granular analysis of the optimal hyperparameters reveals a compelling narrative about how the models adapted to the unique characteristics of each lake and data strategy. While some parameters demonstrated remarkable stability, others were highly sensitive, acting as indicators of data complexity and the efficacy of different training techniques.

Among the most stable hyperparameters were the learning rate (q) and weight decay (r). Across nearly every model, lake, and data configuration, the optimal values were consistently 0.001 and 0.0001, respectively. This stability suggests these settings represent a robust baseline for this specific remote sensing task using Adam-based optimisers. The rare deviations from this standard were strongly correlated with performance degradation. For instance, the pre-trained ResNet50 model for Lake Azul's small patch (Table 5.1) used a learning rate of 0.0001, an order of magnitude smaller, and resulted in a very poor F1-score of 0.689. This reinforces the idea that while tuning is crucial, specific foundational parameters can have a universal optimal range.

In stark contrast, parameters governing model capacity and the learning process, such as batch size (a) and activation function (h), showed significant variability. Models trained on Lake Fogo, which was identified as the more challenging dataset, consistently gravitated towards a smaller batch size of 2 (e.g., see Table 5.3). This preference for smaller batches suggests that more frequent, stochastic gradient updates helped the models navigate a more complex, less uniform loss landscape. Conversely, pre-trained models across both lakes often achieved their best results with a larger batch size of 8. This could be because the pre-learned features provide a much better starting point, allowing the model to fine-tune effectively with more stable gradient estimates from larger batches.

The choice of activation function (h) also revealed dataset-specific adaptations, particularly in response to data augmentation. While ReLU, for CNNs and GELU, for transformers, were the dominant choices, augmentation often enabled other functions to become optimal. For instance, when trained with augmentation on the full image of Lake Fogo (Table 5.4), the best-performing UNet model utilised the tanh function, and DenseNet models often switched to softmax. This shift implies that data augmentation alters the feature distribution in a way that the saturating, zero-centred nature of tanh or the probabilistic outputs of softmax in an intermediate layer can improve performance.

Finally, the transformer architectures provided the most unmistakable evidence of structural adaptation. For both Lake Azul and Lake Fogo, there was a consistent trend: the optimal patch size (k) increased as we moved from the small, localised patch to the whole lake image. This demonstrates an intelligent model adaptation to the scale of the input data; a larger patch size allows the model to capture broader spatial contexts and relationships, which are more prevalent and important when analysing an entire lake.

Concurrently, other transformer-specific parameters, such as the MLP ratio (p), often changed with data augmentation, suggesting an interplay between the model's capacity and the need for regularisation when presented with a richer, more varied dataset. These dynamic adjustments underscore that hyperparameter tuning is not a simple optimisation task but a diagnostic process that reflects deep interactions among model architecture, training strategy, and the intrinsic properties of the underlying data.

Chapter 6

Conclusions & Future Work

This master's dissertation set out to investigate the potential of DL methods for classifying water quality in lakes, leveraging a multimodal approach that combined in-situ measurements with RS imagery. The motivation for this research was the urgent need for reliable, scalable, and automated environmental monitoring systems, particularly for sensitive aquatic ecosystems such as the lakes of São Miguel in the Azores. The core of this work involved a comprehensive evaluation and comparison of various state-of-the-art DL architectures, including CNNs, such as DenseNet and ResNet, and transformer-based models, such as ViT and Swin.

The study rigorously examined the performance of these models under three distinct training paradigms: training from scratch without data augmentation, training with a comprehensive set of data augmentation techniques, and training with pre-trained models (also called TL). For all of the above, 2 distinct datasets were used: one with only a small patch of the sample location, and the second with the full lake's image. By analysing key performance metrics such as F1-score, accuracy, recall, precision, and the often-overlooked but practically critical metric of training time, this dissertation aimed to assess the strengths and limitations of each approach. This final chapter summarises the study in section 6.1 and highlights the main findings in section 6.2. It then revisits the initial research questions in section 6.3, followed by the discussion of the future work in section 6.4. At the end, in section 6.5, closing remarks will be given for this dissertation.

6.1 Summary of the Study

The foundation of this research was a multi-modal dataset constructed from two primary sources, including in-situ measurements of key water quality parameters, including DO, nutrients, BOD, COD, PH, and Chl-a; and multispectral RS imagery from the Copernicus Sentinel-2 satellite mission, retrieved programmatically via the Copernicus API. The data collected spanned 2017-2020, enabling robust temporal alignment between the ground-truth in-situ observations and the corresponding satellite imagery.

A complete and meticulous preprocessing pipeline was designed and implemented for both data modalities. For the tabular dataset of in-situ measurements, missing values were handled using a combination of moving averages and interpolation techniques to preserve temporal integrity. A crucial feature engineering step involved calculating a WQI, which was derived using ecological rules and classification thresholds set by local environmental authorities. This index provided a correct, true label for the classification task based on local regulations.

For the satellite imagery, the pipeline included automated cloud masking and image reconstruction to mitigate the effects of atmospheric interference, a common challenge in the Azores. Following this, various spectral indices with known correlations with water quality, NDCI, NDWI, and turbidity indices, were computed and saved as new images with richer information. To address the quite limited dataset, extensive data augmentation strategies were employed, including geometric transformations (e.g., flipping and affine transformations) and photometric adjustments (e.g., brightness and contrast).

Several prominent DL architectures were implemented from scratch, ensuring full support for both RGB and greyscale images. The models were systematically trained and evaluated under different experimental conditions, for instance, with and without data augmentation, and with the use of pre-trained models from RS datasets. These were also employed using different spatial inputs: a patch-based in-situ collection site and the full image of the lake. A comprehensive grid search over critical hyperparameters was conducted to ensure a fair and extensive exploration of each model's potential.

6.2 Main Findings and Discussion

The experimental results yielded several profound insights into the application of DL for environmental RS. These findings not only answer the core research questions but also develop new ideas for future work in this domain.

The most compelling and unequivocal finding of this research is the transformative effect of data augmentation on model performance. Across all tested architectures, models trained with augmentation strategies consistently and significantly outperformed their non-augmented counterparts. Augmented models achieved the highest F1-score, frequently surpassing 0.979, a substantial improvement over models trained on the original dataset alone. This demonstrates that for a specialised task like water quality classification, which often suffers from a scarcity of diverse, labelled data, techniques such as image rotation, flipping, and colour jittering are not merely supplementary but are, in fact, essential for achieving state-of-the-art results.

Augmentation effectively forces the model to learn more robust and generalizable features by preventing it from overfitting to superficial characteristics of the training images. Furthermore, it was observed that augmentation leads to faster, more stable convergence, requiring fewer training epochs to reach a high-performance plateau. This insight has powerful implications: if a target F1-score can be reached in fewer iterations, it liberates computational resources, allowing researchers to explore more "risky"or complex hyperparameters, such as novel activation functions or larger network architectures, without incurring prohibitive training times.

The study revealed the unpredictable and highly context-dependent nature of transfer learning. The use of pre-trained models yielded mixed results, confirming that this strategy is best described as high-risk, high-reward. On one hand, the pre-trained DenseNet model achieved an exceptional F1-score of 0.981 with a dramatically reduced training time of approximately 100 seconds, demonstrating the immense potential of leveraging knowledge from a source domain. On the other hand, the pre-trained ResNet50 model performed poorly, with its F1-score plummeting to 0.689.

This stark contrast underscores a critical lesson: the success of transfer learning is fundamentally contingent upon the alignment between the features learned by the pre-trained model and the specific characteristics of the target dataset. It is not a universally applicable approach but rather a powerful tool that should be applied strategically, particularly when the pre-trained models are well-suited to the new domain.

Despite inherent challenges posed by persistent cloud coverage and occasional gaps in in-situ data, the research conclusively demonstrated that DL-based classification of the WQI is not only feasible but also highly effective. The models successfully learned to identify the complex spatial and spectral patterns indicative of different water quality levels, capturing temporal trends and ecological dynamics.

While this study showed that DL models are powerful enough to extract meaningful features directly from raw spectral bands, the inclusion of pre-computed spectral indices such as NDCI, NDWI, and others proved invaluable. These indices enrich the input data by providing a mask for the ROI, specifically for the water body and ecologically relevant features, explicitly highlighting properties such as Chl-a concentration and turbidity. Their inclusion made the classification task more robust and provided the models with more direct, interpretable features, likely accelerating the learning process and improving overall performance.

6.3 Revisiting Research Questions and Objectives

Based on the findings and results, this dissertation provides definitive answers to the research questions posed at the outset and confirms the achievement of its primary objectives.

6.3.1 Research Questions

RQ1: To what extent can freely available satellite RS provide accurate, low-cost, and scalable indicators of lake water quality compared with traditional in-situ monitoring? The study demonstrates that freely available Sentinel-2 imagery can provide an accurate, low-cost and scalable tool for routine lake water-quality monitoring. Automated retrieval and preprocessing of Sentinel-2 products, and spectral index creation, enabled monthly monitoring across both case-study lakes, as described in Chapter 3. Classification results and spatial maps show that remote sensing captures large-scale pollution patterns and seasonal trends that align with the in-situ WQI labels, while acknowledging limitations from cloud cover and non-optical parameters. Details on data collection, data visualisation, and data preparation are provided in Sections 3.1, 3.2, and 3.3, respectively. The comparative results appear in Chapter 5 and the finding is summarised in Section 5.3. RQ1 was achieved.

RQ2: Which DL architectures and input representations deliver the best trade-off between classification performance and computational cost when classifying lake water quality? The experiments show a clear trade-off, where transformer-based models and some deeper CNNs achieved the highest classification scores on both the small-patch and full-lake datasets, while lighter CNNs and UNet variants provided competitive accuracy at substantially lower training time. Data augmentation and transfer learning strongly influenced this balance (sometimes dramatically improving accuracy or cutting training time). The experimental setup explanation and the DL conceptions are in Sections 4.1 and 4.2. At the same time, per-architecture performance and measured training times are reported in Chapter 5. RQ2 was achieved.

RQ3: How do the spectral indices help to establish a link with the in-situ data that corresponds to different characteristics of the pollution in water? Spectral indices were explicitly calculated from Sentinel-2 bands. These indices both improve model robustness by emphasising ecologically-relevant signals (Chl-a, turbidity, floating algae) and act as interpretable features that can be corresponded with specific in-situ parameters. The used spectral indices are highlighted in Section 3.3.2. RQ3 was achieved.

6.3.2 Achievement of Objectives

This subsection summarises, objective-by-objective, how the dissertation met the aims listed in Section 1.2.

The first objective was to do an extensive search for the state of the art in water pollution monitoring. RS and IS, specifically Image Classification, were conducted, and 5 related articles were critically reviewed in Sections 2.6 and 2.7. The literature review grounded model choices (CNNs vs Transformers), TL considerations, and the selection of spectral indices used later in the framework. This provided the theoretical basis used throughout the experiments.

The next objective was achieved by constructing a multimodal dataset described in Chapter 3. In-situ measurements were collected and temporally aligned with Sentinel-2 imagery for 2 Azorean lakes; weather data were also integrated. Missing values were handled using interpolation/moving averages. A WQI label was engineered using ecological rules and local thresholds, and spectral indices were computed and stored as model inputs. The data-preparation pipeline is described in Section 3.3.

For the third objective, multiple architectures were implemented and trained, with full support for RGB and greyscale inputs. Several approaches were applied, with and without data augmentation and the use of pre-trained models, as described in Sections 4.1 and 4.2. The models were evaluated on both a local patch (around the sampling site) and full-lake images; per-architecture results and per-lake performance are presented in Chapter 5. Model implementations and experiments demonstrated that both CNNs and transformer-based models can successfully classify WQI categories from Sentinel-2 inputs with supervision from in-situ data.

The fourth objective focused on applying a systematic grid search strategy, as explained in Section 4.1. Hyperparameter spaces for each architecture were explored, and the best candidate models were saved. The grid-search results were logged and analysed to reveal interactions between architecture, approaches, and optimal hyperparameters. This systematic exploration produced the candidate configurations reported in the results tables of Chapter 5.

The last objective was to do an extensive analysis and discussion of the obtained results in Chapter 5. Results show consistent benefits from data augmentation and, in many cases, TL also performed well when the target data matched the original data. Performance metrics (F1, accuracy, recall, precision) and training time were jointly considered to assess trade-offs: transformer and deeper CNN models often achieved the highest F1 scores. At the same time, lighter CNNs and UNet variants delivered competitive accuracy at lower compute cost. The highest reported F1 scores approached 0.9, demonstrating that the framework can produce operationally functional classifications. The discussion also identifies limitations such as cloud gaps and non-optical parameters and proposes mitigation strategies. In Section 6.2, the main findings of this dissertation are discussed.

6.4 Future Work

The conclusions drawn from this study open up numerous exciting avenues for future research. The following directions build directly upon the foundation established in this dissertation and aim to address its limitations.

Given the proven success of data augmentation in accelerating convergence and improving generalisation, all future training should incorporate it as a standard practice. The efficiency gains from this approach should be leveraged to conduct more ambitious experiments. This includes exploring a wider range of hyperparameters, testing more complex model architectures, and experimenting with unconventional loss functions or optimisation algorithms, which would otherwise be computationally prohibitive.

This study focused on a selection of well-established models. Future work should expand this analysis to include more recent and potentially more powerful architectures. This could involve testing advanced variants of the ViT and Swin families, or exploring entirely new paradigms such as diffusion models for data generation, Graph Neural Networkss (GNNs) to model spatial relationships between different parts of a lake, or hybrid CNN-Transformer models that combine the spatial feature extraction power of CNNs with the global context understanding of transformers.

While this study utilised image data augmentation, the tabular in-situ data remained static. A significant area for future work is the development of methods to augment this tabular data synthetically. Techniques like Synthetic Minority Over-sampling Technique (SMOTE) could be used to generate new samples for under-represented classes, such as ”Excellent”or ”Bad”water quality. Furthermore, advanced generative models such as Generative Adversarial Networks (GANs) can be trained to generate new, realistic time-series data points, thereby increasing dataset size and diversity.

A way to increase the confidence and interpretability of the classifications given by the DL models, the usage of XAI provides such interpretability and explanation of how the model made such a classification. Popular libraries, including Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanation (LIME), offer tools to highlight the critical areas of the image where the model looked most closely to achieve this classification.

A highly impactful application of this research would be to focus on point-source pollution, such as the effluent outflows from WWTPs. By using high-resolution satellite imagery to analyse the mixing zone where treated water enters a lake or river, it would be possible to directly monitor the impact of WWTPs on local water quality. This would provide invaluable data for environmental regulation and could help in quickly identifying non-compliant or failing treatment facilities.

The current study utilised Sentinel-2 data with a native spatial resolution of 10m/px. Future research could achieve a more granular analysis by applying super-resolution techniques. DL-based super-resolution models, for example, using GANs, could be used to upscale imagery to 1m/px resolution. This enhanced spatial detail would allow the models to detect finer patterns of pollution, such as small algal blooms or narrow pollution plumes, leading to more accurate and precise classification maps.

A key limitation of relying on a single satellite mission is the vulnerability to data gaps caused by cloud cover or long revisit times. A robust future strategy would involve data fusion, combining imagery from multiple satellite sources such as Sentinel-2 and Landsat. This would create a more complete and temporally dense dataset, improving redundancy and enabling a more continuous analysis of long-term trends, seasonal variations, and short-term pollution events.

To test the models' generalisability, the study area should be expanded to include lakes across different geographical and climatic regions, especially those with less frequent cloud cover. Additionally, the monitoring frequency should be increased from monthly to weekly or bi-weekly. This would allow the detection of ephemeral events, such as harmful algal blooms, which can develop and dissipate rapidly and are often missed by monthly sampling regimes.

6.5 Closing Remarks

This master's dissertation has successfully demonstrated that DL, when combined with data augmentation, transfer learning, and carefully engineered spectral indices, offers a robust, scalable, and cost-effective framework for monitoring water quality. By effectively bridging the gap between RS data and ecological indicators, this research has confirmed that all research questions were answered and all primary objectives were met.

Looking ahead, continuous improvements in satellite technology, data availability, and model architectures will only strengthen this framework. The avenues for future work outlined here represent a clear path toward developing more accurate, frequent, and comprehensive monitoring systems. Ultimately, such advances will provide critical support for sustainable water resource management in the Azores and beyond, contributing to the preservation of our planet's most vital ecosystems.

Beyond the technical contributions, this work reinforces the importance of environmental sustainability as a guiding principle in modern research. Sustainable management of aquatic systems is not only a scientific challenge but also a societal necessity, as clean and accessible water underpins biodiversity, human health, and economic development. By demonstrating how advanced AI and RS can support environmentally responsible decision-making, this dissertation contributes to the broader global agenda of fostering resilience in natural systems while balancing ecological, social, and economic priorities.

Finally, I have also contributed to broader research efforts in environmental monitoring and sustainability. In particular, participation in a paper submitted to the Journal of Applied Logics (IfCoLog) Journal of Logics, titled "Smart Urban Pollution Management Using Deep Learning Techniques for NO_2 Forecasting", highlights a continued commitment to applying DL to pressing environmental challenges beyond water quality. In addition, ongoing interdisciplinary work has led to the submission of another article to a prestigious A* conference, further demonstrating the relevance and impact of AI-driven approaches across multiple domains. This interdisciplinary engagement illustrates the potential of AI to not only monitor but also to anticipate and manage pollution in both natural and urban contexts, further aligning with the global pursuit of sustainable development.

Bibliography

[1] E. E. B. Adam. “Deep learning based NLP techniques in text to speech synthesis for communication recognition.” In: *Journal of Soft Computing Paradigm (JSCP)* 2.04 (2020), pp. 209–215. doi: 10.36548/jscp.2020.4.002.

[2] E. E. Agency. *Water consumption*. accessed: 23/01/2025. url: https://www.eea.europa.eu/en/analysis/publications/the-european-environment-agency-in-brief/inside-the-eea/water-consumption-1?activeTab=658e2886-cfbf-4c2f-a603-061e1627a515.

[3] M. A. S. Al Husaini, M. H. Habaebi, T. S. Gunawan, M. R. Islam, E. A. Elsheikh, and F. Suliman. “Thermal-based early breast cancer detection using inception V3, inception V4 and modified inception MV4.” In: *Neural Computing and Applications* 34.1 (2022), pp. 333–348. doi: 10.1007/s00521-021-06372-1.

[4] M. Altalhan, A. Algarni, and M. Turki-Hadj Alouane. “Imbalanced Data Problem in Machine Learning: A Review.” In: *IEEE Access* 13 (2025), pp. 13686–13699. doi: 10.1109/ACCESS.2025.3531662.

[5] R. Amireddy. “Water Quality Prediction Using Combined Model of Convolutional Neural Network and Long Short-Term Memory.” In: *International Journal of Intelligent Systems and Applications in Engineering* 12.4 (July 2024), pp. 1828–1836. url: https://www.ijisae.org/index.php/IJISAE/article/view/6502.

[6] F. Arena, M. Collotta, L. Luca, M. Ruggieri, and F. G. Termine. “Predictive Maintenance in the Automotive Sector: A Literature Review.” In: *Mathematical and Computational Applications* 27.1 (2022). issn: 2297-8747. doi: 10.3390/mca27010002. url: https://www.mdpi.com/2297-8747/27/1/2.

[7] A. Ashok, H. P. Rani, and K. Jayakumar. “Monitoring of dynamic wetland changes using NDVI and NDWI based landsat imagery.” In: *Remote Sensing Applications: Society and Environment* 23 (2021), p. 100547. issn: 2352-9385. doi: 10.1016/j.rsase.2021.100547. url: https://www.sciencedirect.com/science/article/pii/S2352938521000835.

[8] L. Bailey. accessed: 23/01/2025. url: https://populationeducation.org/what-are-the-three-es-of-sustainability/.

[9] R. Bali Swain and F. Yang-Wallentin. "Achieving sustainable development goals: predicaments and strategies." In: *International Journal of Sustainable Development & World Ecology* 27.2 (2020), pp. 96–106. doi: 10.1080/13504509.2019.1692316.

[10] A. Barbosa-Vasconcelos, Â. Mendes, F. Martins, E. Lopes, A. Machado, A. A. Bordalo, P. Vaz-Pires, N. Vieira, P. Martins da Costa, and L. J. Bessa. "River water analysis using a multiparametric approach: Portuguese river as a case study." In: *Journal of Water and Health* 16.6 (2018), pp. 991–1006. doi: 10.2166/wh.2018.047.

[11] A. G. Barto. "Reinforcement Learning: An Introduction." In: *SIAM Rev* 6.2 (2021), p. 423. doi: 10.1016/S0893-6080(99)00098-2.

[12] J. Botero-Valencia, C. Barrantes-Toro, D. Marquez-Viloria, and J. M. Pearce. "Low-cost air, noise, and light pollution measuring station with wireless communication and tinyML." In: *HardwareX* 16 (2023), e00477. issn: 2468-0672. doi: 10.1016/j.ohx.2023.e00477. url: https://www.sciencedirect.com/science/article/pii/S2468067223000846.

[13] R. Boucekkine, G. Fabbri, S. Federico, and F. Gozzi. "From firm to global-level pollution control: The case of transboundary pollution." In: *European Journal of Operational Research* 290.1 (2021), pp. 331–345. issn: 0377-2217. doi: 10.1016/j.ejor.2020.07.056. url: https://www.sciencedirect.com/science/article/pii/S0377221720306822.

[14] J.-P. Boussemart, H. Leleu, Z. Shen, and V. Valdmanis. "Performance analysis for three pillars of sustainability." In: *Journal of Productivity Analysis* 53.3 (2020), pp. 305–320. doi: 10.1007/s11123-020-00575-9.

[15] R. Calegari, G. Ciatto, V. Mascardi, and A. Omicini. "Logic-based technologies for multi-agent systems: a systematic literature review." In: *Autonomous Agents and Multi-Agent Systems* 35.1 (2021), p. 1. doi: 10.1007/s10458-020-09478-3.

[16] R. de Carvalho Bueno. "Wind-induced internal waves in closed basins: A laboratory experiment and field study." Doctoral dissertation. Federal University of Paraná, Feb. 2019. doi: 10.13140/RG.2.2.35265.20328.

[17] R. Cerveny. "The history of the world meteorological organization." In: *Weatherwise* 76.2 (2023), pp. 34–40. doi: 10.1080/00431672.2023.2159222.

[18] J. Chaki and N. Dey. *A beginner's guide to image preprocessing techniques*. CRC Press, 2018. doi: 10.1201/9780429441134.

[19] F. N. Chaudhry and M. Malik. "Factors affecting water pollution: a review." In: *J. Ecosyst. Ecography* 7.1 (2017), pp. 225–231. doi: 10.4172/2157-7625.1000225.

[20] P. Chetty. *Satellite technology and its applications*. Notion Press, 2023. isbn: 978-1-118-63647-3.

[21] K. N. Clasen, L. Hackel, T. Burgert, G. Sumbul, B. Demir, and V. Markl. *reBEN: Refined BigEarthNet Dataset for Remote Sensing Image Analysis*. 2025. doi: 10.48550/arXiv.2407.03653. arXiv: 2407.03653 [cs.CV]. url: https://arxiv.org/abs/2407.03653.

[22] M. Claverie, E. F. Vermote, B. Franch, and J. G. Masek. "Evaluation of the Landsat-5 TM and Landsat-7 ETM+ surface reflectance products." In: *Remote Sensing of Environment* 169 (2015), pp. 390–403. issn: 0034-4257. doi: 10.1016/j.rse.2015.08.030. url: https://www.sciencedirect.com/science/article/pii/S0034425715301188.

[23] S. Clerc, C. Donlon, F. Borde, N. Lamquin, S. E. Hunt, D. Smith, M. McMillan, J. Mittaz, E. Woolliams, M. Hammond, C. Banks, T. Moreau, B. Picard, M. Raynal, P. Rieu, and A. Guérou. "Benefits and Lessons Learned from the Sentinel-3 Tandem Phase." In: *Remote Sensing* 12.17 (2020). issn: 2072-4292. doi: 10.3390/rs12172668. url: https://www.mdpi.com/2072-4292/12/17/2668.

[24] G. B. Courrèges-Lacoste, M. Sallusti, G. Bulsa, G. Bagnasco, B. Veihelmann, S. Riedl, D. J. Smith, and R. Maurer. "The Copernicus Sentinel 4 mission: a geostationary imaging UVN spectrometer for air quality monitoring." In: *Sensors, Systems, and Next-Generation Satellites XXI*. Ed. by S. P. Neeck, J.-L. Bézy, and T. Kimura. Vol. 10423. International Society for Optics and Photonics. SPIE, 2017, p. 1042307. doi: 10.1117/12.2282158. url: https://doi.org/10.1117/12.2282158.

[25] S. Cruz, M. Ahmed Ouameur, and F. Pereira de Figueiredo. "Reinforcement Learning-based Wi-Fi Contention Window Optimization." In: *Journal of Communication and Information Systems* 38 (Jan. 2023), pp. 128–143. doi: 10.14209/jcis.2023.15.

[26] N. Dabire, E. C. Ezin, and A. M. Firmin. "Water Quality Assessment Using Normalized Difference Index by Applying Remote Sensing Techniques: Case of Lake Nokoue." In: *2024 IEEE 15th Control and System Graduate Research Colloquium (ICSGRC)*. 2024, pp. 1–6. doi: 10.1109/ICSGRC62081.2024.10690936.

[27] M. Danilevsky, K. Qian, R. Aharonov, Y. Katsis, B. Kawas, and P. Sen. *A Survey of the State of Explainable AI for Natural Language Processing*. 2020. doi: 10.48550/arXiv.2010.00711. arXiv: 2010.00711 [cs.CL]. url: https://arxiv.org/abs/2010.00711.

[28] T. Davie. *Fundamentals of hydrology*. Routledge, 2019. doi: 10.4324/9780203798942.

[29] C. J. F. Delcourt, A. Combee, B. Izbicki, M. C. Mack, T. Maximov, R. Petrov, B. M. Rogers, R. C. Scholten, T. A. Shestakova, D. van Wees, and S. Veraverbeke. "Evaluating the Differenced Normalized Burn Ratio for Assessing Fire Severity Using Sentinel-2 Imagery in Northeast Siberian Larch Forests." In: *Remote Sensing* 13.12 (2021). issn: 2072-4292. doi: 10.3390/rs13122311. url: https://www.mdpi.com/2072-4292/13/12/2311.

[30] C. W. E. C. Department. accessed: 21/01/2025. url: https://esamultimedia.esa.int/multimedia/publications/ESABulletin161/offline/download.pdf.

[31] C. J. Donlon, R. Cullen, L. Giulicchi, P. Vuilleumier, C. R. Francis, M. Kuschnerus, W. Simpson, A. Bouridah, M. Caleno, R. Bertoni, J. Rancaño, E. Pourier, A. Hyslop, J. Mulcahy, R. Knockaert, C. Hunter, A. Webb, M. Fornari, P. Vaze, S. Brown, J. Willis, S. Desai, J.-D. Desjonqueres, R. Scharroo, C. Martin-Puig, E. Leuliette, A. Egido, W. H. Smith, P. Bonnefond, S. Le Gac, N. Picot, and G. Tavernier. "The Copernicus Sentinel-6 mission: Enhanced continuity of satellite sea level measurements from space." In: *Remote Sensing of Environment* 258 (2021), p. 112395. issn: 0034-4257. doi: 10.1016/j.rse.2021.112395. url: https://www.sciencedirect.com/science/article/pii/S0034425721001139.

[32] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. 2021. doi: 10.48550/arXiv.2010.11929. arXiv: 2010.11929 [cs.CV]. url: https://arxiv.org/abs/2010.11929.

[33] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, et al. "Explainable AI (XAI): Core ideas, techniques, and solutions." In: *ACM computing surveys* 55.9 (2023), pp. 1–33. doi: 10.1145/3561048.

[34] A. V. Egorov, D. P. Roy, H. K. Zhang, Z. Li, L. Yan, and H. Huang. "Landsat 4, 5 and 7 (1982 to 2017) Analysis Ready Data (ARD) Observation Coverage over the Conterminous United States and Implications for Terrestrial Monitoring." In: *Remote Sensing* 11.4 (2019). issn: 2072-4292. doi: 10.3390/rs11040447. url: https://www.mdpi.com/2072-4292/11/4/447.

[35] C. Elachi and J. J. Van Zyl. *Introduction to the physics and techniques of remote sensing*. John Wiley & Sons, 2021. doi: 10.5670/oceanog.2006.39.

[36] "Enhanced scaling effects significantly lower the ability of MODIS normalized difference snow index to estimate fractional and binary snow cover on the Tibetan Plateau." In: *Journal of Hydrology* 592 (2021), p. 125795. issn: 0022-1694. doi: 10.1016/j.jhydrol.2020.125795.

[37] N. Eramma, H. Lalita, S. Satishgouda, S. R. Jyothi, C. Venkatesh, and S. J. Patil. "Zooplankton Productivity Evaluation of Lentic and Lotic Ecosystem." In: *Limnology*. Ed. by C. Massarelli and C. Campanale. Rijeka: IntechOpen, 2023. Chap. 1. doi: 10.5772/intechopen.107020.

[38] W. Ertel. *Introduction to artificial intelligence*. Springer Nature, 2024. doi: 10.1007/978-3-658-43102-0.

[39] A. E. Evans, J. Mateo-Sagasta, M. Qadir, E. Boelee, and A. Ippolito. "Agricultural water pollution: key knowledge gaps and research needs." In: *Current Opinion in Environmental Sustainability* 36 (2019). Environmental Change Assessment, pp. 20–27. issn: 1877-3435. doi: 10.1016/j.cosust.2018.10.003. url: https://www.sciencedirect.com/science/article/pii/S1877343518300873.

[40] R. FALKNER. "The Paris Agreement and the new logic of international climate politics." In: *International Affairs* 92.5 (Aug. 2016), pp. 1107–1125. issn: 0020-5850. doi: 10.1111/1468-2346.12708. eprint: https://academic.oup.com/ia/article-pdf/92/5/1107/13653532/inta12708.pdf. url: 10.1111/1468-2346.12708.

[41] M. Geissdoerfer, P. Savaget, N. M. Bocken, and E. J. Hultink. "The Circular Economy – A new sustainability paradigm?" In: *Journal of Cleaner Production* 143 (2017), pp. 757–768. issn: 0959-6526. doi: 10.1016/j.jclepro.2016.12.048. url: https://www.sciencedirect.com/science/article/pii/S0959652616321023.

[42] S. E. Gibbs, C. P. Salgado Kent, B. Slat, D. Morales, L. Fouda, and J. Reisser. "Cetacean sightings within the great Pacific Garbage Patch." In: *Marine Biodiversity* 49.4 (2019), pp. 2021–2027. doi: 10.1007/s12526-019-00952-0.

[43] E. Golovina and K. Karennik. "Modern Trends in the Field of Solving Transboundary Problems in Groundwater Extraction." In: *Resources* 10.10 (2021). issn: 2079-9276. doi: 10.3390/resources10100107. url: https://www.mdpi.com/2079-9276/10/10/107.

[44] M. Gomez, J. Perdiguero, and A. Sanz. "Socioeconomic factors affecting water access in rural areas of low and middle income countries." In: *Water* 11.2 (2019), p. 202. doi: 10.3390/w11020202.

[45] R. Gou, R. Shi, Q. Zhang, G. Yang, Z. Wang, H.-L. Zheng, and X. Tuo. "ResNeXt Deep Learning Model-Based Transmission Image Reconstruction of Tomographic Gamma Scanning With Array Detectors." In: *IEEE Transactions on Nuclear Science* 72.1 (2025), pp. 61–72. doi: 10.1109/TNS.2024.3511550.

[46] A. Government. *Plano de Gestão da Região Hidrográfica dos Açores (PGRH-Açores)*. Accessed on: 2025/05/17. 2021. url: https://www.azores.gov.pt/Gra/srrn-drotrh/conteudos/livres/PGRH-A%C3%A7ores.htm.

[47] M. Gridach, J. Nanavati, K. Zine El Abidine, L. Mendes, and C. Mack. "Agentic AI for Scientific Discovery: A Survey of Progress, Challenges, and Future Directions." In: *arXiv preprint arXiv:2503.08979* (2025). Preprint, submitted March 12, 2025.

[48] B. Grizzetti, O. Vigiak, A. Udias, A. Aloe, M. Zanni, F. Bouraoui, A. Pistocchi, C. Dorati, R. Friedland, A. De Roo, C. Benitez Sanz, A. Leip, and M. Bielza. "How EU policies could reduce nutrient pollution in European inland and coastal waters." In: *Global Environmental Change* 69 (2021), p. 102281. issn: 0959-3780. doi: 10.1016/j.gloenvcha.2021.102281. url: https://www.sciencedirect.com/science/article/pii/S0959378021000601.

[49] A. Guerrisi, I. Falcone, F. Valenti, M. Rao, E. Gallo, S. Ungania, M. Maccallini, M. Fanciulli, P. Frascione, A. Morrone, and M. Caterino. "Artificial Intelligence and Advanced Melanoma: Treatment Management Implications." In: *Cells* 11 (Dec. 2022), p. 3965. doi: 10.3390/cells11243965.

[50] J. Gupta, S. Pathak, and G. Kumar. "Deep learning (CNN) and transfer learning: a review." In: *Journal of Physics: Conference Series*. Vol. 2273. 1. IOP Publishing. 2022, p. 012029. doi: 10.1088/1742-6596/2273/1/012029.

[51] L. Hackel, K. N. Clasen, and B. Demir. "ConfigILM: A general purpose configurable library for combining image and language models for visual question answering." In: *SoftwareX* 26 (2024), p. 101731. issn: 2352-7110. doi: 10.1016/j.softx.2024.101731. url: https://www.sciencedirect.com/science/article/pii/S235271102400102X.

[52] D. Hanelli, A. Barth, G. Volkmer, and M. Köhler. "Modelling of Acid Mine Drainage in Open Pit Lakes Using Sentinel-2 Time-Series: A Case Study from Lusatia, Germany." In: *Minerals* 13 (Feb. 2023). doi: 10.3390/min13020271.

[53] P. R. Hannah Ritchie and M. Roser. accessed: 23/01/2025. url: https://ourworldindata.org/co2-and-greenhouse-gas-emissions.

[54] P. R. Hannah Ritchie and M. Roser. accessed: 23/01/2025. url: https://ourworldindata.org/co2-and-greenhouse-gas-emissions.

[55] B. M. Hashim, M. A. Sultan, M. N. Attyia, A. A. Al Maliki, and N. Al-Ansari. "Change Detection and Impact of Climate Changes to Iraqi Southern Marshes Using Landsat 2 MSS, Landsat 8 OLI and Sentinel 2 MSI Data and GIS Applications." In: *Applied Sciences* 9.10 (2019). issn: 2076-3417. doi: 10.3390/app9102016. url: https://www.mdpi.com/2076-3417/9/10/2016.

[56] S. H. Hassan Al-Taai. "Water pollution Its causes and effects." In: *IOP Conference Series: Earth and Environmental Science* 790.1 (June 2021), p. 012026. doi: 10.1088/1755-1315/790/1/012026. url: https://dx.doi.org/10.1088/1755-1315/790/1/012026.

[57] K. He, X. Zhang, S. Ren, and J. Sun. *Deep Residual Learning for Image Recognition*. 2015. doi: 10.48550/arXiv.1512.03385. arXiv: 1512.03385 [cs.CV]. url: https://arxiv.org/abs/1512.03385.

[58] R. A. Hill, C. C. Moore, J. M. Doyle, S. G. Leibowitz, P. L. Ringold, and B. Rashleigh. "The Clean Water Act After 50 Years: Innovations in Measuring the Social Benefits of Water Quality for Research and Policy: Estimating biotic integrity to capture existence value of freshwater ecosystems." In: *Proceedings of the National Academy of Sciences of the United States of America* 120.18 (2023).

[59] C.-C. Ho, M. A. Benalcazar Hernandez, Y.-F. Chen, C.-J. Lin, and C.-S. Chen. "Deep Residual Neural Network-Based Defect Detection on Complex Backgrounds." In: *IEEE Transactions on Instrumentation and Measurement* 71 (2022), pp. 1–10. doi: 10.1109/TIM.2022.3144224.

[60] S. D. Holcomb, W. K. Porter, S. V. Ault, G. Mao, and J. Wang. "Overview on DeepMind and Its AlphaGo Zero AI." In: *Proceedings of the 2018 International Conference on Big Data and Education*. ICBDE '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 67–71. isbn: 9781450363587. doi: 10.1145/3206157.3206174. url: https://doi.org/10.1145/3206157.3206174.

[61] A. A. Hopgood. *Intelligent systems for engineers and scientists: a practical guide to artificial intelligence*. CRC press, 2021. doi: 10.1201/9781003226277.

[62] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. *Densely Connected Convolutional Networks*. 2018. doi: 10.48550/arXiv.1608.06993. arXiv: 1608.06993 [cs.CV]. url: https://arxiv.org/abs/1608.06993.

[63] M. Iman, H. R. Arabnia, and K. Rasheed. "A Review of Deep Transfer Learning and Recent Advancements." In: *Technologies* 11.2 (2023). issn: 2227-7080. doi: 10.3390/technologies11020040. url: https://www.mdpi.com/2227-7080/11/2/40.

[64] S. Ioffe and C. Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. doi: 10.48550/arXiv.1502.03167. arXiv: 1502.03167 [cs.LG]. url: https://arxiv.org/abs/1502.03167.

[65] J. Irizar, M. Melf, P. Bartsch, J. Koehler, S. Weiss, R. Greinacher, M. Erdmann, V. Kirschner, A. P. Albinana, and D. Martin. "Sentinel-5/UVNS." In: *International Conference on Space Optics — ICSO 2018*. Ed. by Z. Sodnik, N. Karafolas, and B. Cugny. Vol. 11180. International Society for Optics and Photonics. SPIE, 2019, p. 1118004. doi: 10.1117/12.2535923. url: https://doi.org/10.1117/12.2535923.

[66] A. Isen, M. Rossin-Slater, and W. R. Walker. "Every breath you take—every dollar you'll make: The long-term consequences of the clean air act of 1970." In: *Journal of Political Economy* 125.3 (2017), pp. 848–902. doi: 10.1086/691465.

[67] C. Janiesch, P. Zschech, and K. Heinrich. "Machine learning and deep learning." In: *Electronic Markets* 31.3 (2021), pp. 685–695. doi: 10.1007/s12525-021-00475-2.

[68] B. K. Jha, S. G. G, and V. K. R. "E-Commerce Product Image Classification using Transfer Learning." In: *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. 2021, pp. 904–912. doi: 10.1109/ICCMC51019.2021.9418371.

[69] B. Jiang, A. E. Kauffman, L. Li, W. McFee, B. Cai, J. Weinstein, J. R. Lead, S. Chatterjee, G. I. Scott, and S. Xiao. "Health impacts of environmental contamination of micro-and nanoplastics: a review." In: *Environmental health and preventive medicine* 25.1 (2020), p. 29. doi: 10.1186/s12199-020-00870-9.

[70] D. Jin, N. Gunner, N. Carvajal Janke, S. Baruah, K. M. Gold, and Y. Jiang. "Aleks: AI powered Multi-Agent System for Autonomous Scientific Discovery via Data-Driven Approaches in Plant Science." In: *arXiv preprint arXiv:2508.19383* (2025). Preprint, submitted August 26, 2025.

[71] S. Jutz and M. Milagro-Pérez. "Copernicus: the European Earth Observation programme." In: *Revista de Teledetección* 56 (Nov. 2020), pp. V–XI. doi: 10.4995/raet.2020.14346. url: https://polipapers.upv.es/index.php/raet/article/view/14346.

[72] M. D. N. Kamlesh Lulla. "Mission to earth: LANDSAT 9 will continue to view the world." In: *Geocarto International* 36.20 (2021), pp. 2261–2263. doi: 10.1080/10106049.2021.1991634.

[73] M. Khan, A. Hanan, M. Kenzhebay, M. Gazzea, and R. Arghandeh. "Transformer-based land use and land cover classification with explainability using satellite imagery." In: *Scientific Reports* 14.1 (2024), p. 16744. doi: 10.1038/s41598-024-67186-4.

[74] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet classification with deep convolutional neural networks." In: *Commun. ACM* 60.6 (May 2017), pp. 84–90. issn: 0001-0782. doi: 10.1145/3065386. url: https://doi.org/10.1145/3065386.

[75] D. P. Lam, B. Martín-López, A. Wiek, E. M. Bennett, N. Frantzeskaki, A. I. Horcea-Milcu, and D. J. Lang. "Scaling the impact of sustainability initiatives: a typology of amplification processes." In: *Urban Transformations* 2.1 (2020), p. 3. doi: 10.1186/s42854-020-00007-9.

[76] P. Lemos, P. Silva, C. A. Sousa, and A. J. Duarte. "Polluted Rivers—A Case Study in Porto, Portugal." In: *Ecologies* 5.2 (2024), pp. 188–197. issn: 2673-4133. doi: 10.3390/ecologies5020012. url: https://www.mdpi.com/2673-4133/5/2/12.

[77] Y.-H. Li, Y.-L. Li, M.-Y. Wei, and G.-Y. Li. "Innovation and challenges of artificial intelligence technology in personalized healthcare." In: *Scientific reports* 14.1 (2024), p. 18994. doi: 10.1038/s41598-024-70073-7.

[78] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou. "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects." In: *IEEE Transactions on Neural Networks and Learning Systems* 33.12 (2022), pp. 6999–7019. doi: 10.1109/TNNLS.2021.3084827.

[79] Q. Lian, L. Yao, Z. Uddin Ahmad, X. Lei, F. Islam, M. E. Zappi, and D. D. Gang. "Nonpoint source pollution." In: *Water Environment Research* 91.10 (2019), pp. 1114–1128. doi: 10.1002/wer.1205. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wer.1205. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/wer.1205.

[80] T. Lin, Y. Wang, X. Liu, and X. Qiu. "A survey of transformers." In: *AI Open* 3 (2022), pp. 111–132. issn: 2666-6510. doi: 10.1016/j.aiopen.2022.10.001. url: https://www.sciencedirect.com/science/article/pii/S2666651022000146.

[81] A. Liu, Y. Ma, J. M. Gunawardena, P. Egodawatta, G. A. Ayoko, and A. Goonetilleke. "Heavy metals transport pathways: The importance of atmospheric pollution contributing to stormwater pollution." In: *Ecotoxicology and Environmental Safety* 164 (2018), pp. 696–703. issn: 0147-6513. doi: 10.1016/j.ecoenv.2018.08.072. url: https://www.sciencedirect.com/science/article/pii/S014765131830811X.

[82] Y. Liu, B. Su, H. Mu, Y. Zhang, L. Chen, and B. Wu. "Effects of point and nonpoint source pollution on urban rivers: From the perspective of pollutant composition and toxicity." In: *Journal of Hazardous Materials* 460 (2023), p. 132441. doi: 10.1016/j.jhazmat.2023.132441.

[83] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*. 2021. doi: 10.48550/arXiv.2103.14030. arXiv: 2103.14030 [cs.CV]. url: https://arxiv.org/abs/2103.14030.

[84] T. R. Loveland and J. R. Irons. "Landsat 8: The plans, the reality, and the legacy." In: *Remote Sensing of Environment* 185 (2016). Landsat 8 Science Results, pp. 1–6. issn: 0034-4257. doi: 10.1016/j.rse.2016.07.033. url: https://www.sciencedirect.com/science/article/pii/S0034425716302905.

[85] J. Lu. "Turnover of environmental protection officials and transboundary water pollution." In: *Environmental Science and Pollution Research* 28.8 (2021), pp. 10207–10223. doi: 10.1007/s11356-020-11530-4.

[86] I. P. Macário, B. B. Castro, I. M. Nunes, C. Pizarro, C. Coelho, F. Gonçalves, and D. R. de Figueiredo. "Stepwise strategy for monitoring toxic cyanobacterial blooms in lentic water bodies." In: *Environmental monitoring and assessment* 189 (2017), pp. 1–11. doi: 10.1007/s10661-017-6292-9.

[87] M. S. Mancini, A. Galli, V. Niccolucci, D. Lin, S. Bastianoni, M. Wackernagel, and N. Marchettini. "Ecological Footprint: Refining the carbon Footprint calculation." In: *Ecological Indicators* 61 (2016), pp. 390–403. issn: 1470-160X. doi: 10.1016/j.ecolind.2015.09.040. url: https://www.sciencedirect.com/science/article/pii/S1470160X15005269.

[88] T. Mannucci and E.-J. van Kampen. "A hierarchical maze navigation algorithm with Reinforcement Learning and mapping." In: *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2016, pp. 1–8. doi: 10.1109/SSCI.2016.7849365.

[89] J. M. S. Martínez. "The Spanish-Portuguese Transboundary Waters Agreements: Historic Perspective." In: *Water International* 28.3 (2003), pp. 379–388. doi: 10.1080/02508060308691712.

[90] A. J. Meddens, C. A. Kolden, and J. A. Lutz. "Detecting unburned areas within wildfire perimeters using Landsat and ancillary data across the northwestern United States." In: *Remote Sensing of Environment* 186 (2016), pp. 275–285. issn: 0034-4257. doi: 10.1016/j.rse.2016.08.023. url: https://www.sciencedirect.com/science/article/pii/S0034425716303261.

[91] H. Meng, J. Zhang, Z. Zheng, Y. Song, and Y. Lai. "Classification of inland lake water quality levels based on Sentinel-2 images using convolutional neural networks and spatiotemporal variation and driving factors of algal bloom." In: *Ecological Informatics* 80 (2024), p. 102549. issn: 1574-9541. doi: https://doi.org/10.1016/j.ecoinf.2024.102549. url: https://www.sciencedirect.com/science/article/pii/S1574954124000918.

[92] M. Mezzacapo, K. Falinski, C. Smith, A. El-Kadi, M. Donohue, and D. Lerner. "Hawai'i's Cesspool Problem: Review and Recommendations for Water Resources and Human Health." In: *Journal of Contemporary Water Research Education* 170 (Nov. 2020). doi: 10.1111/j.1936-704X.2020.03339.x.

[93] L. Missions. accessed: 21/01/2025. url: https://www.usgs.gov/media/images/spectral-bandpasses-all-landsat-sensors.

[94] N. Morin-Crini, E. Lichtfouse, G. Liu, V. Balaram, A. R. L. Ribeiro, Z. Lu, F. Stock, E. Carmona, M. R. Teixeira, L. A. Picos-Corrales, et al. "Worldwide cases of water pollution by emerging contaminants: a review." In: *Environmental Chemistry Letters* 20.4 (2022), pp. 2311–2338. doi: 10.1007/s10311-022-01447-4.

[95] E. Mousavinasab, N. Zarifsanaiey, S. R. Niakan Kalhori, M. Rakhshan, L. Keikha, and M. Ghazi Saeedi. "Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods." In: *Interactive Learning Environments* 29.1 (2021), pp. 142–163. doi: 10.1080/10494820.2018.1558257.

[96] D. A. F. Muzhoffar, Y. Sakuno, N. Taniguchi, K. Hamada, H. Shimabukuro, and M. Hori. "Automatic Detection of Floating Macroalgae via Adaptive Thresholding Using Sentinel-2 Satellite Data with 10 m Spatial Resolution." In: *Remote Sensing* 15.8 (2023). issn: 2072-4292. doi: 10.3390/rs15082039. url: https://www.mdpi.com/2072-4292/15/8/2039.

[97] S. Naeem, A. Ali, S. Anam, and M. M. Ahmed. "An unsupervised machine learning algorithms: Comprehensive review." In: *International Journal of Computing and Digital Systems* (2023). doi: 10.12785/ijcds/130172.

[98] V. Nasteski. "An overview of the supervised machine learning methods." In: *Horizons. b* 4.51-62 (2017), p. 56. doi: 10.20544/HORIZONS.B.04.1.17.P05.

[99] J Nevado, R. Martín-Doimeadios, F. Guzmán Bernardo, M. Jiménez-Moreno, S Tardío, M Fornieles, S Ríos, and A Pérez. "Integrated pollution evaluation of the Tagus River in Central Spain." In: *Environmental monitoring and assessment* 156 (Sept. 2008), pp. 461–77. doi: 10.1007/s10661-008-0498-9.

[100] M. for the Environment of New Zealand. accessed: 23/01/2025. url: https://environment.govt.nz/publications/our-freshwater-2020/issue-2-water-is-polluted-in-urban-farming-and-forestry-areas/.

[101] Y. Ning, S. He, Z. Wu, C. Xing, and L.-J. Zhang. "A Review of Deep Learning Based Speech Synthesis." In: *Applied Sciences* 9.19 (2019). issn: 2076-3417. doi: 10.3390/app9194050. url: https://www.mdpi.com/2076-3417/9/19/4050.

[102] P. Oliveira, B. Fernandes, F. Aguiar, M. A. Pereira, and P. Novais. "Evaluating Unidimensional Convolutional Neural Networks to Forecast the Influent pH of Wastewater Treatment Plants." In: *Intelligent Data Engineering and Automated Learning – IDEAL 2021*. Cham: Springer International Publishing, 2021, pp. 446–457. isbn: 978-3-030-91608-4. doi: 10.1007/978-3-030-91608-4_44.

[103] P. Oliveira, F. S. Marcondes, M. S. Duarte, D. Durães, C. Gonçalves, G. Martins, and P. Novais. "Incdualpathnet: a hybrid architecture proposal for predicting energy production in a wastewater treatment plants." In: *Neural Computing and Applications* (2025), pp. 1–29. doi: 10.1007/s00521-025-11545-3.

[104] O. Özhan. "Short-time-Fourier transform." In: *Basic transforms for electrical engineering*. Springer, 2022, pp. 441–464. doi: 10.1007/978-3-030-98846-3_7.

[105] J. B. Palter. "The Role of the Gulf Stream in European Climate." In: *Annual Review of Marine Science* 7.Volume 7, 2015 (2015), pp. 113–137. issn: 1941-0611. doi: 10.1146/annurev-marine-010814-015656. url: https://www.annualreviews.org/content/journals/10.1146/annurev-marine-010814-015656.

[106] A. S. Paymode and V. B. Malode. "Transfer Learning for Multi-Crop Leaf Disease Image Classification using Convolutional Neural Network VGG." In: *Artificial Intelligence in Agriculture* 6 (2022), pp. 23–33. issn: 2589-7217. doi: 10.1016/j.aiia.2021.12.002. url: https://www.sciencedirect.com/science/article/pii/S2589721721000416.

[107] P. Potin, O. Colin, M. Pinheiro, B. Rosich, A. O'Connell, T. Ormston, J.-B. Gratadour, and R. Torres. "Status And Evolution Of The Sentinel-1 Mission." In: *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*. 2022, pp. 4707–4710. doi: 10.1109/IGARSS46834.2022.9884753.

[108] F. Pu, C. Ding, Z. Chao, Y. Yu, and X. Xu. "Water-Quality Classification of Inland Lakes Using Landsat8 Images by Convolutional Neural Networks." In: *Remote Sensing* 11.14 (2019). issn: 2072-4292. doi: 10.3390/rs11141674. url: https://www.mdpi.com/2072-4292/11/14/1674.

[109] B. Purvis, Y. Mao, and D. Robinson. "Three pillars of sustainability: in search of conceptual origins." In: *Sustainability science* 14 (2019), pp. 681–695. doi: 10.1007/s11625-018-0627-5.

[110] N. N. Rabalais and R. E. Turner. "Gulf of Mexico hypoxia: Past, present, and future." In: *Limnology and Oceanography Bulletin* 28.4 (2019), pp. 117–124. doi: 10.1002/lob.10351.

[111] K. Raiyani, T. Gonçalves, L. Rato, P. Salgueiro, and J. R. Marques da Silva. "Sentinel-2 Image Scene Classification: A Comparison between Sen2Cor and a Machine Learning Approach." In: *Remote Sensing* 13.2 (2021). issn: 2072-4292. doi: 10.3390/rs13020300. url: https://www.mdpi.com/2072-4292/13/2/300.

[112] F. Y. Ramírez-Castillo, A. Loera-Muro, M. Jacques, P. Garneau, F. J. Avelar-González, J. Harel, and A. L. Guerrero-Barrera. "Waterborne Pathogens: Detection Methods and Challenges." In: *Pathogens* 4.2 (2015), pp. 307–334. issn: 2076-0817. doi: 10.3390/pathogens4020307. url: https://www.mdpi.com/2076-0817/4/2/307.

[113] A. R. Reshi, S. Pichuka, and A. Tripathi. "Applications of Sentinel-5P TROPOMI Satellite Sensor: A Review." In: *IEEE Sensors Journal* 24.13 (2024), pp. 20312–20321. doi: 10.1109/JSEN.2024.3355714.

[114] A. R. L. Ribeiro, A. Maia, M. Santos, M. Tiritan, and C. Ribeiro. "Occurrence of Natural Contaminants of Emerging Concern in the Douro River Estuary, Portugal." In: *Archives of environmental contamination and toxicology* 70 (Aug. 2015). doi: 10.1007/s00244-015-0212-1.

[115] H. Ritchie, M. Roser, and P. Rosado. "Renewable Energy." In: *Our World in Data* (2020). https://ourworldindata.org/renewable-energy.

[116] C. Rochman, A. Andrady, S. Dudas, J. Fabres, F. Galgani, D. lead, V. Hidalgo-Ruz, S. Hong, P. Kershaw, L. Lebreton, A. Lusher, R. Narayan, S. Pahl, J. Potemra, C. Rochman, S. Sherif, J. Seager, W. Shim, P. Sobral, and L. Amaral-Zettler. *SOURCES, FATE AND EFFECTS OF MICROPLASTICS IN THE MARINE ENVIRONMENT: PART 2 OF A GLOBAL ASSESSMENT*. Dec. 2016.

[117] O. Rochmawanti and F. Utaminingrum. "Chest X-Ray Image to Classify Lung diseases in Different Resolution Size using DenseNet-121 Architectures." In: *Proceedings of the 6th International Conference on Sustainable Information Engineering and Technology*. SIET '21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 327–331. isbn: 9781450384070. doi: 10.1145/3479645.3479667. url: https://doi.org/10.1145/3479645.3479667.

[118] J. Romero-Luis, A. Carbonell-Alcocer, M. Gertrudix, M. D. C. Gertrudis Casado, P. Giardullo, and D. Wuebben. "Recommendations to improve communication effectiveness in social marketing campaigns: Boosting behavior change to foster a circular economy." In: *Cogent Social Sciences* 8.1 (2022), p. 2147265. doi: 10.1080/23311886.2022.2147265.

[119] V. Sagan, K. T. Peterson, M. Maimaitijiang, P. Sidike, J. Sloan, B. A. Greeling, S. Maalouf, and C. Adams. "Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing." In: *Earth-Science Reviews* 205 (2020), p. 103187. issn: 0012-8252. doi: 10.1016/j.earscirev.2020.103187. url: https://www.sciencedirect.com/science/article/pii/S0012825220302336.

[120] V. Sagan, K. T. Peterson, M. Maimaitijiang, P. Sidike, J. Sloan, B. A. Greeling, S. Maalouf, and C. Adams. "Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing." In: *Earth-Science Reviews* 205 (2020), p. 103187. issn: 0012-8252. doi: 10.1016/j.earscirev.2020.103187. url: https://www.sciencedirect.com/science/article/pii/S0012825220302336.

[121] A. Salazar-Torres, D. Robles, A. Reyes, M. C. Santa-Maria, and P. Venail. "Evaluation of planktonic cyanobacteria in Peruvian freshwater lentic water bodies: prevalence and regulatory framework to aid policy making." In: *Environmental Monitoring and Assessment* 195.7 (2023), p. 852. doi: 10.1007/s10661-023-11487-0.

[122] M. Sauerwein, E. Doubrovski, R. Balkenende, and C. Bakker. "Exploring the potential of additive manufacturing for product design in a circular economy." In: *Journal of Cleaner Production* 226 (2019), pp. 1138–1149. doi: 10.1016/j.jclepro.2019.04.108.

[123] H. Schandl. Accessed: 23/01/2025. url: https://ensia.com/voices/economic-development-resource-consumption/.

[124] D. S. Schmeller, A. Loyau, K. Bao, W. Brack, A. Chatzinotas, F. De Vleeschouwer, J. Friesen, L. Gandois, S. V. Hansson, M. Haver, G. Le Roux, J. Shen, R. Teisserenc, and V. T. Vredenburg. "People, pollution and pathogens – Global change impacts in mountain freshwater ecosystems." In: *Science of The Total Environment* 622-623 (2018), pp. 756–763. issn: 0048-9697. doi: 10.1016/j.scitotenv.2017.12.006. url: https://www.sciencedirect.com/science/article/pii/S0048969717334253.

[125] J. Segarra, M. L. Buchaillot, J. L. Araus, and S. C. Kefauver. "Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications." In: *Agronomy* 10.5 (2020). issn: 2073-4395. doi: 10.3390/agronomy10050641. url: https://www.mdpi.com/2073-4395/10/5/641.

[126] E. P. R. Service. accessed: 23/01/2025. url: https://www.europarl.europa.eu/topics/en/article/20151201STO05603/circular-economy-definition-importance-and-benefits.

[127] R. Shah, R. Chen, and M. Woydt. "The effects of energy efficiency and resource consumption on environmental sustainability." In: *Lubricants* 9.12 (2021), p. 117. doi: 10.3390/lubricants9120117.

[128] N. K. Shrestha, R. P. Rudra, P. Daggupati, P. K. Goel, and R. Shukla. "A comparative evaluation of the continuous and event-based modelling approaches for identifying critical source areas for sediment and phosphorus losses." In: *Journal of Environmental Management* 277 (2021), p. 111427. issn: 0301-4797. doi: 10.1016/j.jenvman.2020.111427. url: https://www.sciencedirect.com/science/article/pii/S0301479720313529.

[129] A. L. Silva, P. Oliveira, D. Durães, D. Fernandes, R. Névoa, J. Monteiro, P. Melo-Pinto, J. Machado, and P. Novais. "A Framework for Representing, Building and Reusing Novel State-of-the-Art Three-Dimensional Object Detection Models in Point Clouds Targeting Self-Driving Applications." In: *Sensors* 23.14 (2023). issn: 1424-8220. doi: 10.3390/s23146427. url: https://www.mdpi.com/1424-8220/23/14/6427.

[130] K. Simonyan and A. Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. doi: 10.48550/arXiv.1409.1556. arXiv: 1409.1556 [cs.CV]. url: https://arxiv.org/abs/1409.1556.

[131] R. K. Sinha and B. G. Loganathan. "Ganges river contamination: a review." In: *Water Challenges and Solutions on a Global Scale* (2015), pp. 129–159. doi: 10.1021/bk-2015-1206.ch008.

[132] A. J. Stewart, C. Robinson, I. A. Corley, A. Ortiz, J. M. Lavista Ferres, and A. Banerjee. "TorchGeo: Deep Learning With Geospatial Data." In: *ACM Transactions on Spatial Algorithms and Systems* (Dec. 2024). doi: 10.1145/3707459. url: https://doi.org/10.1145/3707459.

[133] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. "Inception-v4, inception-resnet and the impact of residual connections on learning." In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 31. 1. 2017. doi: 10.1609/aaai.v31i1.11231.

[134] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. *Going Deeper with Convolutions*. 2014. doi: 10.48550/arXiv.1409.4842. arXiv: 1409.4842 [cs.CV]. url: https://arxiv.org/abs/1409.4842.

[135] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. *Rethinking the Inception Architecture for Computer Vision*. 2015. doi: 10.48550/arXiv.1512.00567. arXiv: 1512.00567 [cs.CV]. url: https://arxiv.org/abs/1512.00567.

[136] L. A. Thompson and W. S. Darwish. "Environmental chemical contaminants in food: review of a global problem." In: *Journal of toxicology* 2019.1 (2019), p. 2345283. doi: 10.1155/2019/2345283.

[137] S. Thylashri, A. Manikanta, I. Poojitha, and K. Varshitha. "Water Quality Monitoring with Satellite Image Using Convolutional Neural Network." In: *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. 2023, pp. 1–5. doi: 10.1109/ICCCNT56998.2023.10306619.

[138] J.-N. Thépaut, D. Dee, R. Engelen, and B. Pinty. "The Copernicus Programme and its Climate Change Service." In: *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*. 2018, pp. 1591–1593. doi: 10.1109/IGARSS.2018.8518067.

[139] E. S. A. E. Union. *Copernicus Browser*. Accessed on: 2025/07/14. 2025. url: https://browser.dataspace.copernicus.eu/.

[140] Z. Wang, J. Huo, and Y. Duan. "The impact of government incentives and penalties on willingness to recycle plastic waste: An evolutionary game theory perspective." In: *Frontiers of Environmental Science & Engineering* 14.2 (2020), p. 29. doi: 10.1007/s11783-019-1208-2.

[141] S. B. Watson, C. Miller, G. Arhonditsis, G. L. Boyer, W. Carmichael, M. N. Charlton, R. Confesor, D. C. Depew, T. O. Höök, S. A. Ludsin, G. Matisoff, S. P. McElmurry, M. W. Murray, R. Peter Richards, Y. R. Rao, M. M. Steffen, and S. W. Wilhelm. "The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia." In: *Harmful Algae* 56 (2016), pp. 44–66. issn: 1568-9883. doi: 10.1016/j.hal.2016.04.010. url: https://www.sciencedirect.com/science/article/pii/S1568988315301141.

[142] M. A. Wulder, T. R. Loveland, D. P. Roy, C. J. Crawford, J. G. Masek, C. E. Woodcock, R. G. Allen, M. C. Anderson, A. S. Belward, W. B. Cohen, J. Dwyer, A. Erb, F. Gao, P. Griffiths, D. Helder, T. Hermosilla, J. D. Hipple, P. Hostert, M. J. Hughes, J. Huntington, D. M. Johnson, R. Kennedy, A. Kilic, Z. Li, L. Lymburner, J. McCorkel, N. Pahlevan, T. A. Scambos, C. Schaaf, J. R. Schott, Y. Sheng, J. Storey, E. Vermote, J. Vogelmann, J. C. White, R. H. Wynne, and Z. Zhu. "Current status of Landsat program, science, and applications." In: *Remote Sensing of Environment* 225 (2019), pp. 127–147. issn: 0034-4257. doi: 10.1016/j.rse.2019.02.015. url: https://www.sciencedirect.com/science/article/pii/S0034425719300707.

[143] M. A. Wulder, D. P. Roy, V. C. Radeloff, T. R. Loveland, M. C. Anderson, D. M. Johnson, S. Healey, Z. Zhu, T. A. Scambos, N. Pahlevan, M. Hansen, N. Gorelick, C. J. Crawford, J. G. Masek, T. Hermosilla, J. C. White, A. S. Belward, C. Schaaf, C. E. Woodcock, J. L. Huntington, L. Lymburner, P. Hostert, F. Gao, A. Lyapustin, J.-F. Pekel, P. Strobl, and B. D. Cook. "Fifty years of Landsat science and impacts." In: *Remote Sensing of Environment* 280 (2022), p. 113195. issn: 0034-4257. doi: 10.1016/j.rse.2022.113195. url: https://www.sciencedirect.com/science/article/pii/S0034425722003054.

[144] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He. *Aggregated Residual Transformations for Deep Neural Networks*. 2017. doi: 10.48550/arXiv.1611.05431. arXiv: 1611.05431 [cs.CV]. url: https://arxiv.org/abs/1611.05431.

[145] Y. Yamada, R. T. Lange, C. Lu, S. Hu, C. Lu, J. Foerster, and D. Ha. "The AI Scientist-v2: Workshop-Level Automated Scientific Discovery via Agentic Tree Search." In: *arXiv preprint arXiv:2504.08066* (2025). Preprint, submitted April 10, 2025.

[146] D. D. Yavaşlı, C. J. Tucker, and K. A. Melocik. "Change in the glacier extent in Turkey during the Landsat Era." In: *Remote Sensing of Environment* 163 (2015), pp. 32–41. issn: 0034-4257. doi: 10.1016/j.rse.2015.03.002. url: https://www.sciencedirect.com/science/article/pii/S003442571500098X.

[147] K. Yoshino, K. Yamada, M. Tanaka, Y. Tada, G. Kanaya, Y. Henmi, and M. Yamamoto. "Subtidal benthic communities in Minamata Bay, Japan, approximately 30 years after mercury pollution remediation involving dredging disturbance." In: *Ecological Research* 37.1 (2022), pp. 137–150. doi: 10.1111/1440-1703.12281. eprint: https://esj-journals.onlinelibrary.wiley.com/doi/pdf/10.1111/1440-1703.12281. url: https://esj-journals.onlinelibrary.wiley.com/doi/abs/10.1111/1440-1703.12281.

[148] H. Zarini, A. Khalili, H. Tabassum, M. Rasti, and W. Saad. "AlexNet Classifier and Support Vector Regressor for Scheduling and Power Control in Multimedia Heterogeneous Networks." In: *IEEE Transactions on Mobile Computing* 22.5 (2023), pp. 2520–2536. doi: 10.1109/TMC.2021.3123200.

[149] Z. Zhu, Y. Zhou, K. C. Seto, E. C. Stokes, C. Deng, S. T. Pickett, and H. Taubenböck. "Understanding an urbanizing planet: Strategic directions for remote sensing." In: *Remote Sensing of Environment* 228 (2019), pp. 164–182. issn: 0034-4257. doi: `10.1016/j.rse.2019.04.020`. url: `https://www.sciencedirect.com/science/article/pii/S0034425719301658`.

[150] H. Üzen, M. Türkoğlu, B. Yanikoglu, and D. Hanbay. "Swin-MFINet: Swin transformer based multi-feature integration network for detection of pixel-level surface defects." In: *Expert Systems with Applications* 209 (2022), p. 118269. issn: 0957-4174. doi: `10.1016/j.eswa.2022.118269`. url: `https://www.sciencedirect.com/science/article/pii/S0957417422014105`.