BULGARIAN ACADEMY OF SCIENCES
INSTITUTE OF MATHEMATICS AND INFORMATICS

DOCTORAL THESIS

Adaptive Neural Network for Processing Satellite Data with Different Spatial and Spectral Characteristics

---

Author:
Ventsislav POLIMENOV

Supervisor:
Krassimira IVANOVA
Assoc. Prof. PhD

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

Field of Higher Education 4. Natural Sciences, Mathematics and Informatics
Professional Field 4.6. Informatics
Doctoral Programme "Informatics"

April 30, 2026

Declaration of Authorship

I, Ventsislav POLIMENOV, declare that this thesis titled, “Adaptive Neural Network for Processing Satellite Data with Different Spatial and Spectral Characteristics” and the work presented in it are my own. I confirm that:

- This work was completed during my PhD study at the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences.
- Any part of this thesis has not previously been submitted for a degree or any other qualification.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. Except for such quotations, this thesis is entirely my work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

To my wife

For the support, the trust and for being there through all of it. This would not have been possible without you.

To my children

May this be a small reminder that ideas alone are weightless. It is consistency, perseverance and the courage to keep going that transform them into something real.

To my mother and grandmother

For the unconditional love and unwavering belief that I could do this, even in the moments I could not believe it myself.

Ventsislav POLIMENOV

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Krassimira Ivanova, for her guidance, patience and the many fruitful conversations that steered this work toward its completion. Her support throughout the PhD process has been invaluable.

I am deeply grateful to Kamen Iliev and the RST-TTO team for the initial push that set this work in motion, for the ideas, data and opportunities generously shared along the way and for the sustained intellectual engagement that shaped so much of what is presented here. I would also like to thank Ivelin Ivanov for the helpful discussions around LAI construction and target data and for the explanations and suggestions that meaningfully informed the methodology.

The computational experiments presented in this thesis were carried out on the HEMUS high-performance computing cluster, and I am grateful for the access and support provided.

Finally, I owe a quiet but profound debt of gratitude to my family, whose unconditional support and trust carried me through the more demanding stretches of this journey.

BULGARIAN ACADEMY OF SCIENCES
INSTITUTE OF MATHEMATICS AND INFORMATICS

Abstract

Adaptive Neural Network for Processing Satellite Data with Different Spatial and Spectral Characteristics

by Ventsislav POLIMENOV

A thesis submitted in fulfillment of the requirements for the degree Doctor of Philosophy

Leaf Area Index (LAI), defined as the one-sided green leaf area per unit ground surface area, is a fundamental biophysical parameter for quantifying vegetation structure, health and productivity. Accurate LAI estimation is critical for precision agriculture, enabling crop monitoring, resource optimisation and yield prediction. Traditional field-based methods provide high accuracy but are labour-intensive and spatially limited, rendering them impractical for landscape-scale monitoring. Satellite remote sensing offers a scalable alternative, yet existing approaches face persistent limitations – empirical vegetation indices saturate at moderate-to-high LAI values, physics-based radiative transfer model inversions are computationally demanding and require site-specific parameterisation and most methods process each sensor independently, failing to exploit complementary information across the growing constellation of Earth observation satellites.

This dissertation develops a multi-sensor deep learning framework for operational LAI estimation using freely available Sentinel-2 and Landsat 8/9 imagery over agricultural regions in Bulgaria. The framework centres on a **U-Net convolutional neural network with Conditional Batch Normalization and Atrous Spatial Pyramid Pooling (ASPP)**, which maintains sensor-specific normalization parameters while sharing convolutional weights to learn sensor-invariant vegetation representations. Spatially dense training labels are generated through a vegetation index ensemble approach that averages three Beer-Lambert-transformed indices (NDVI, GNDVI, SAVI), producing sensor-agnostic pseudo-LAI targets that enable supervised training without field measurements. A batch balancing strategy using 4× differential augmentation sampling of the minority sensor addresses the ~4:1 Landsat-to-Sentinel-2 tile imbalance, preventing catastrophic mode collapse during multi-sensor training. Validation employs spatial block cross-validation ensuring that all tiles within the same geographic block are assigned to the same split, eliminating spatial data leakage.

The multi-sensor model achieves $R^2 = 0.999$ with RMSE of 0.017 m^2/m^2 under spatially rigorous validation, representing 80% and 64% RMSE reductions for Sentinel-2 and Landsat 8/9 respectively compared to single-sensor baselines. Ablation analysis reveals that batch balancing accounts for approximately 85% of the multi-sensor performance improvement, reframing the problem as primarily one of data distribution rather than architectural complexity. Geographic generalisation is demonstrated through zero-shot transfer to the Burgas region, located 200 km east of the Plovdiv training site, where both sensors achieve RMSE of 0.051 m^2/m^2 without retraining. Temporal stability is confirmed across an 8-month gap using May 2024 imagery, with the model producing biologically plausible LAI increases consistent with spring vegetation growth. The CNN achieves 77% lower RMSE than the best individual vegetation index, demonstrating that spatial learning from neighbouring pixels extracts information beyond what pixel-level spectral indices can resolve.

The limitations of the VI-ensemble pseudo-label approach are transparently acknowledged – the CNN is trained and primarily evaluated against targets derived from the same spectral bands it receives as input. The work is therefore framed as a methodological contribution to multi-sensor remote sensing and a VI-ensemble enhancement tool, rather than a claim of field-validated absolute LAI accuracy. The framework establishes a foundation for future extension toward field-calibrated training, additional biophysical indices and integration of very high resolution commercial satellite imagery.

Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
ANNSIA	Adaptive Neural Network for Satellite Image Analysis
ASPP	Atrous Spatial Pyramid Pooling
BOA	Bottom Of Atmosphere
BRDF	Bidirectional Reflectance Distribution Function
CAVIS	Clouds, Aerosols, Vapours, Ice and Snow system
CCCI	Canopy Chlorophyll Content Index
CGLS	Copernicus Global Land Service
CI	Chlorophyll Index
CNN	Convolutional Neural Network
CondBN	Conditional Batch Normalization
CRS	Coordinate Reference System
ESA	European Space Agency
EVI	Enhanced Vegetation Index
EWT	Equivalent Water Thickness
GNDVI	Green Normalized Difference Vegetation Index
GNSS	Global Navigation Satellite System
GWC	Gravimetric Water Content
HEMUS	Abbreviation for HEterogeneous MUlti-functional Supercomputer
HPC	High Performance Computing
HS	Hyperspectral
L8 / L9	Landsat 8 / Landsat 9
LAI	Leaf Area Index
LiDAR	Light Detection and Ranging
LWC	Leaf Water Content
MAE	Mean Absolute Error
ML	Machine Learning

MS	Multispectral
MSI	Multispectral Instrument (Sentinel-2)
NaN	Not a Number
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near Infrared
OLI	Operational Land Imager
PAN	Panchromatic
PROSAIL	Combines PROSPECT leaf model and SAIL canopy model
PROSPECT	PROjection of leaf SPECtral Transmittance and reflectance
QA_PIXEL	Quality Assessment Pixel Band
RMSE	Root Mean Square Error
RTM	Radiative Transfer Model
RWC	Relative Water Content
S2	Sentinel-2
SAIL	Scattering by Arbitrary Inclined Leaves
SAVI	Soil Adjusted Vegetation Index
SCL	Scene Classification Layer
SNAP	Sentinel Application Platform
SWIR	Shortwave Infrared
TIRS	Thermal Infrared Sensor
TOA	Top Of Atmosphere
UAV	Unmanned Aerial Vehicle
USGS	United States Geological Survey
UV	Ultraviolet
UVA / UVB / UVC	UV radiation types
VI	Vegetation Index
VNIR	Visible and Near Infrared

Chapter 1

Introduction

Leaf Area Index (LAI), defined as the one-sided green leaf area per unit ground surface area, is a fundamental biophysical parameter for quantifying vegetation structure, health and productivity [34]. Accurate LAI estimation is critical for precision agriculture, enabling farmers and agronomists to monitor crop development, optimize irrigation and fertilization, detect plant stress and predict yields. Traditional field-based LAI measurement methods – such as destructive sampling or hemispherical photography – are labour-intensive, time-consuming and spatially limited, making them impractical for large-scale agricultural monitoring [56].

Satellite remote sensing offers a scalable alternative, providing spatially continuous LAI estimates across extensive agricultural landscapes at regular temporal intervals. However, existing approaches face several challenges. Empirical vegetation index methods often have limited sensitivity and tend to saturate at high LAI values, while physics-based radiative transfer approaches are computationally demanding and require site-specific calibration [109]. Most critically, the challenge of integrating multi-sensor satellite data – which differ in spatial resolution, spectral characteristics and temporal coverage – remains largely unsolved, as current methods typically process each sensor independently and fail to exploit complementary information across sensors.

This thesis addresses these challenges by developing a multi-sensor deep learning framework for operational LAI estimation using freely available Sentinel-2 and Landsat 8/9 satellite imagery. The approach combines convolutional neural networks with novel training strategies to overcome the lack of field-measured ground truth, achieve robust cross-sensor generalization and demonstrate unprecedented accuracy in agricultural LAI mapping.

1.1 Purpose

Over the past few decades, advancements in satellite technology and machine learning algorithms have revolutionized the way we monitor and manage agricultural land. The ability to collect and analyse vast amounts of data using satellite imagery and deep learning has opened up new avenues for improving crop yields, managing resources and ensuring food safety [34]. The resulting models are robust, efficient and trustworthy, providing farmers with valuable insights into crop health and productivity. With further research and development, this technology has the potential to revolutionise the agricultural industry, increasing efficiency and sustainability while ensuring the safety and security of our food supply.

This research seeks to explore the use of satellite imagery and machine learning algorithms for smart agriculture and food safety, with a focus on the estimation of LAI using Sentinel-2 and Landsat 8/9 satellite data from agricultural regions in Bulgaria. Through this investigation, we aim to demonstrate the practical applicability of multi-sensor deep learning approaches in operational agricultural monitoring and establish a foundation for future advancements in precision agriculture.

1.2 Motivation

The study examines the latest advancements in both remote sensing and deep learning as applied to LAI estimation. We have identified several critical limitations that must be addressed to advance the capabilities and potential uses of these technologies in operational settings. Developing a unified neural network capable of processing multispectral images across varying spatial resolutions and spectral characteristics would represent a significant advancement in remote sensing technology. Such a network would need to robustly handle the details and complexities of different image types, allowing for more flexible and comprehensive analysis.

Agricultural decision-making increasingly relies on timely and accurate biophysical information, as LAI directly relates to photosynthetic capacity, water use efficiency and biomass accumulation. However, operational LAI products face substantial limitations. While operational satellite products such as MODIS LAI and the Sentinel-2 SNAP Biophysical Processor provide global coverage, they often show large biases or inconsistencies when validated against ground measurements. Empirical vegetation indices are computationally efficient but suffer from saturation at moderate to high LAI values, with NDVI typically plateauing around LAI values of three to five, alongside persistent sensitivity to soil background and sensor-specific calibration requirements [48]. Physics-based radiative transfer model inversion methods such as PROSAIL are theoretically robust but computationally demanding, require prior knowledge of vegetation parameters and often produce ill-posed solutions [109]. Field measurements provide the highest accuracy but are limited to point locations and specific dates, offering insufficient spatial and temporal coverage for landscape-scale applications.

The emergence of deep learning offers a potential breakthrough, as convolutional neural networks can learn complex spectral-spatial LAI relationships directly from data, bypassing the need for explicit physical models while achieving higher accuracy than empirical vegetation indices. However, this approach introduces a fundamental challenge regarding how to obtain spatially dense training labels when field-measured LAI is unavailable at the scale required for training deep neural networks.

Furthermore, our research highlighted the significant differences in spectral bands across various satellite probes and their sensors. Modern earth observation provides multiple complementary data sources, including Sentinel-2 with its ten-meter spatial resolution, ten spectral bands including red-edge channels and five-day revisit time, as well as Landsat 8/9 with thirty-meter spatial resolution, six spectral bands, sixteen-day revisit time and a long-term archive extending back to 1972. When used together, these sensors provide an effective revisit time of two to three days. Despite this data richness, most LAI estimation methods process each sensor independently, treating them as separate problems and missing the fundamental opportunity to develop a single model that can learn to estimate LAI from multiple sensors simultaneously while exploiting their complementary spectral and spatial information.

The technical obstacles to multi-sensor integration are substantial. Processing ten-meter and thirty-meter imagery in a unified architecture requires addressing resolution mismatch, as Sentinel-2 has ten bands including unique red-edge channels while Landsat 8 has only six bands. Unequal tile counts per sensor can cause catastrophic mode collapse during multi-sensor training and ensuring the model produces consistent LAI estimates regardless of input sensor remains a significant challenge. Moving forward, we anticipate conducting follow-up research aimed at developing methods to normalize these spectral variations, facilitating the creation of a unified analytical framework for multispectral image processing.

Another critical challenge, often overlooked in deep learning studies, is spatial data leakage in validation. Agricultural landscapes exhibit strong spatial autocorrelation, as neighbouring pixels share soil properties, crop types and management practices. Random train-test splits commonly used in machine learning risk placing spatially correlated pixels in both training and validation sets, inflating performance estimates and failing to assess true generalization capability. Proper validation for agricultural remote sensing requires geographic separation between training and validation sets, ensuring the model learns transferable spatial patterns rather than memorizing pixel-level correlations. Yet most deep learning studies for biophysical parameter estimation use random splits, making their generalization claims questionable and highlighting the need for more rigorous spatial validation methodologies.

1.3 Goals and Objectives

The primary goal of this thesis is to develop and evaluate an adaptive neural network capable of processing satellite imagery from multiple sources with different spatial and spectral resolutions. The principal application is the estimation of the Leaf Area Index, with potential extension to other biophysical properties such as water content and chlorophyll concentration.

To achieve this overarching goal, we have established several **specific objectives** that guide the research and development process.

The first objective centres on data preprocessing and preparation, where we collect and pre-process satellite imagery from Landsat 8/9, Sentinel-2 to establish a comprehensive dataset. This involves generating a unified dataset from pre-corrected surface reflectance products (Sentinel-2 Level-2 and Landsat Collection 2 Level-2), with consistent tiling across sensors to create a controlled environment that enables systematic comparative experiments across different sensor configurations and resolutions.

The second objective addresses model architecture and implementation. We design and implement a multi-sensor convolutional neural network architecture that can process variable-resolution multispectral data from different sensors within a unified framework. This involves incorporating advanced techniques such as Conditional Batch Normalization to manage sensor-specific spectral characteristics while maintaining a shared representation of vegetation structure. The architecture must handle the challenge of processing Sentinel-2 imagery at ten-meter resolution with ten spectral bands alongside Landsat 8/9 imagery at thirty-meter resolution with six spectral bands, all while producing consistent LAI estimates regardless of the input sensor.

The third objective focuses on the prediction of biophysical properties. We train the model to estimate LAI using multi-source satellite imagery, generating spatially dense training labels through an ensemble vegetation index approach that creates pseudo-LAI targets enabling supervised training without requiring extensive field measurements. This approach involves averaging multiple vegetation index-based estimates to reduce individual biases while maintaining the sensor-agnostic consistency required for effective multi-sensor training. We systematically evaluate the impact of varying spatial and spectral resolutions on prediction accuracy to understand the fundamental trade-offs and opportunities in multi-sensor LAI estimation.

The fourth objective addresses a critical challenge in multi-sensor training through the development of batch balancing strategies. We investigate and resolve the epoch-level exposure imbalance problem that arises when training on datasets with unequal tile counts per sensor. In our Plovdiv study region, Landsat 8 provides 4.2 times more tiles than Sentinel-2 due to its broader scene coverage, which can cause catastrophic mode collapse where one sensor fails while the other converges. Our solution implements differential augmentation sampling, applying higher sampling rates to the minority sensor to equalize epoch-level exposure and prevent the network from becoming biased toward the overrepresented sensor.

The fifth objective establishes rigorous validation methodologies that address the often-overlooked problem of spatial data leakage. We implement spatial block cross-validation using a four-by-four geographic grid partitioning strategy that enforces strict spatial separation of five to ten kilometres between training and validation sets. This ensures honest assessment of generalization capability by preventing spatially autocorrelated pixels from appearing in both training and validation sets, thereby forcing the model to learn transferable spatial patterns rather than memorizing pixel-level correlations.

The sixth objective validates the model's ability to generalize beyond the training domain through comprehensive cross-region and temporal testing. We apply the Plovdiv-trained model to Burgas, a geographically distinct agricultural region located approximately two hundred kilometres away, without any retraining or model modification. Additionally, we assess temporal stability by evaluating the 2023-trained model on May 2024 Burgas imagery, confirming that the learned spectral-LAI relationships remain valid across an eight-month gap despite differences in crop phenology, weather patterns and atmospheric conditions.

The seventh objective quantifies the value added by the convolutional neural network approach compared to traditional methods. We systematically compare the trained model's performance against direct vegetation index computation, measuring improvements over the constituent vegetation indices that serve as training labels. This analysis demonstrates whether spatial learning and multi-sensor fusion can extract more robust LAI estimates from the same spectral information and quantifies the magnitude of improvement achievable through deep learning compared to empirical approaches.

The final objective supports practical deployment and scientific transparency through tool development and documentation. We develop methodologies for operational biophysical index prediction that demonstrate the model's real-world applicability in agricultural monitoring scenarios. Furthermore, we document and make available the complete data processing pipeline, model architecture specifications and evaluation metrics to ensure transparency and enable reproducibility of our findings by the broader research community.

1.4 Hypothesis

Based on these objectives and the current state of remote sensing technology, we formulate three primary hypotheses that guide our experimental design and validation approach.

The first hypothesis suggests that a spectral-aware convolutional neural network can generalize across sensors at least as effectively as sensor-specific models, despite differences in spatial resolution and spectral band configuration. This challenges the conventional wisdom that each sensor requires a dedicated model and suggests that learning sensor-invariant LAI features may actually improve performance compared to sensor-specific representations.

The second hypothesis proposes that spatial resolution affects LAI prediction accuracy non-linearly, with the relationship between resolution and performance depending on the spatial scale of agricultural features and the strength of spatial autocorrelation in the landscape. This suggests that the commonly assumed direct relationship between higher resolution and better performance may not hold universally, particularly when finer resolution introduces additional noise or when the LAI signal is predominantly captured at coarser spatial scales.

The third hypothesis proposes that a model trained on spectral-LAI relationships in one agricultural region can generalise to geographically and temporally distinct regions without retraining or recalibration, provided that predictions are expressed in absolute physical units. Under this hypothesis, cross-region performance degradation in regression models arises primarily from the encoding of training-region distributional statistics into the output representation rather than from architectural limitations or spectral domain shift.

1.5 Dissertation Outline

This dissertation is structured to provide a comprehensive exploration of multi-sensor deep learning for LAI estimation, progressing logically from foundational concepts through methodological development to empirical validation and synthesis of findings.

Chapter 2: Research Background presents a thorough review of the state-of-the-art in satellite remote sensing for LAI estimation. This chapter covers the evolution from traditional vegetation indices through physics-based radiative transfer models to modern deep learning approaches, positioning our work within the broader context of multi-sensor data fusion and biophysical parameter retrieval. We analyse the strengths and limitations of existing methods, identifying critical gaps that motivate our research direction and establishing the theoretical foundation for our methodological choices.

Chapter 3: Methodology details the complete methodology for multi-sensor LAI estimation through an eleven-stage pipeline. The chapter begins with Stage Zero, which addresses LAI ground truth generation via the vegetation index ensemble approach, creating pseudo-labels that enable supervised training without extensive field measurements. Subsequent stages cover data loading, resampling to handle different sensor resolutions and tiling strategies that create consistent geographic coverage across sensors. We then describe dataset generation procedures, augmentation techniques for balancing multi-sensor training data and normalization approaches including our novel Conditional Batch Normalization method. The model architecture section presents the Multi-Sensor U-Net design incorporating Atrous Spatial Pyramid Pooling, followed by detailed exposition of the training strategy emphasizing our breakthrough batch balancing solution. The methodology chapter concludes with comprehensive coverage of evaluation metrics and our rigorous five-tier validation framework encompassing spatial validation through geographic block splitting, comparison with operational products, radiative transfer model validation, cross-sensor consistency checks and empirical vegetation index benchmarking.

Chapter 4: Experiments and Results documents the experimental validation of our approach through comprehensive empirical evaluation, focusing primarily on experiments employing spatially rigorous validation with four-by-four geographic block splitting to eliminate spatial data leakage. We present single-sensor baseline results establishing independent performance benchmarks for Sentinel-2-only and Landsat 8/9-only configurations. The chapter then details Experiment X3, our best multi-sensor model using VI-ensemble targets, which achieves overall R-squared of 0.999 with RMSE of 0.017 square meters per square meter, representing 80% improvement for Sentinel-2 and 64% improvement for Landsat 8/9 compared to their respective single-sensor baselines. We include ablation studies examining hybrid VI-plus-PROSAIL targets, which degraded performance and demonstrated the critical importance of sensor-agnostic ground truth in multi-sensor training. The validation extends beyond the training region through cross-region testing in Burgas, located two hundred kilometres from the Plovdiv training site, where both sensors achieve identical RMSE of 0.051 despite no model retraining. Temporal validation using May 2024 Burgas imagery confirms robust temporal transfer over an eight-month gap, with the model producing biologically plausible LAI increases consistent with spring vegetation growth. The chapter concludes with systematic comparison against operational products including the ESA SNAP Biophysical Processor and quantification of the 77% RMSE improvement our CNN achieves over the best individual vegetation index component.

Chapter 5: Conclusions and Future Work synthesizes the research contributions and contextualizes our findings within the broader scientific landscape. This concluding chapter highlights the five primary contributions of the thesis: a multi-sensor U-Net architecture with Conditional Batch Normalization for heterogeneous sensor fusion, batch balancing as a critical enabler for multi-sensor training convergence, a VI-ensemble pseudo-label approach enabling supervised training without field measurements, a multi-tier validation framework with spatial block cross-validation and cross-region generalization and GPU-accelerated training with full-scene inference on HPC infrastructure. We provide transparent discussion of limitations, particularly the circular ground truth problem inherent in using VI-ensemble pseudo-labels for both training and evaluation, framing our contribution honestly as a methodological advancement in spatial remote sensing and a VI-ensemble enhancement tool rather than claiming validated absolute LAI accuracy. The chapter concludes with proposed future research directions including field validation campaigns to ground-truth the VI-ensemble calibration, extension to cross-season generalization requiring multi-temporal training data and integration of complementary data sources such as Sentinel-1 synthetic aperture radar for cloud-penetrating capabilities.

Chapter 2

Research Background

2.1 Literature Review

The rapid development of technology especially over the past decade has led to a significant leap in the advancement of remote sensing technologies and their use in various disciplines including environmental monitoring, agriculture, urban planning and disaster management. Technological innovations in the field of satellites and unmanned aerial vehicle (UAV) platforms as well as significant improvements in sensors for capturing various terrestrial features have greatly increased the accessibility of remote sensing data and created the possibility of using the collected data to conduct intelligent analysis that support informed decision-making.

Collaborative efforts between academia, industry, governments, and international organisations have further accelerated innovation in remote sensing technologies. These developments have resulted in improved spatial resolution, broader temporal coverage, and more advanced data processing techniques. Of great importance in this process is the integration of remote sensing with the disciplines related to the analysis of large volumes of data and artificial intelligence, especially deep learning. This integration has created new opportunities for extracting meaningful information from large and complex Earth observation datasets.

The increased interest in remote sensing, as well as its integration with the capabilities of artificial intelligence techniques, is also evidenced by the increasing number of articles indexed in Scopus. Figure 2.1 shows this trend based on the number of publications over the last ten years, filtering in keywords by (1) "remote sensing" and (2) "remote sensing" in combination with ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural networks").

Figure 2.2 further demonstrates that research in this domain spans multiple scientific fields. A significant proportion of publications fall within the areas of Computer Science and Mathematics (approximately 25%), followed by Engineering (17%). At the same time, many studies focus on applying computational techniques to support research in Earth and Planetary Sciences, Physics and Astronomy, Environmental Science, and Agricultural and Biological Sciences.

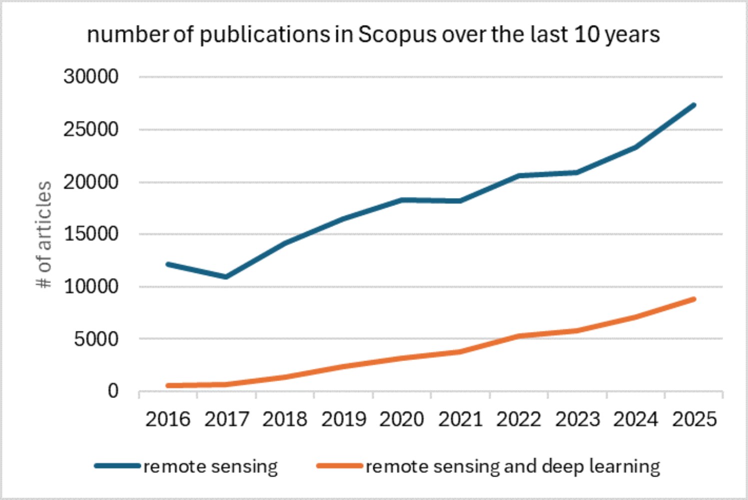


FIGURE 2.1: Number of Scopus-indexed publications on "remote sensing" and "remote sensing combined with AI/ML techniques", 2016–2025.

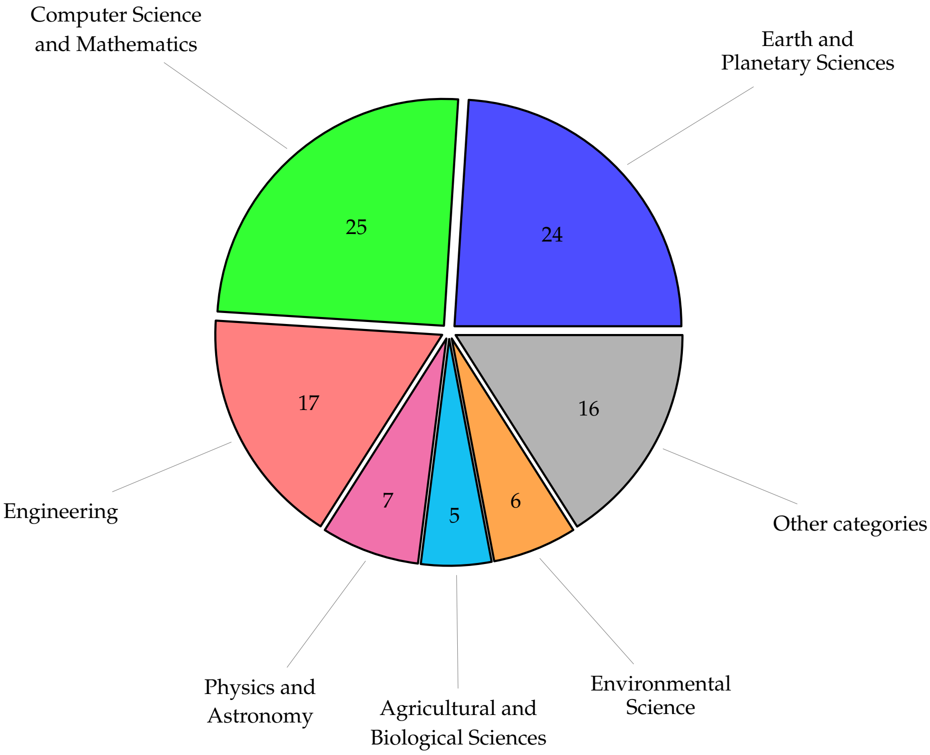


FIGURE 2.2: Distribution of publications across subject areas (with some classified in multiple categories).

Due to the rapidly growing interest in this field, this chapter reviews the current state-of-the-art methods and methodologies for applying deep learning models to remote sensing data [37, 61]. The review also highlights existing gaps and misconceptions in the current understanding and applications of satellite-based Earth observation technologies. Based on this analysis, potential directions for improving accuracy, efficiency, and reliability in remote sensing applications are discussed.

A key focus of this research is the assessment of vegetation-related parameters that are essential for monitoring plant health and ecosystem dynamics. Particular attention is given to biophysical properties that can be extracted or inferred from satellite and UAV imagery. These indicators not only serve to understand plant structure and function, but also support broader goals related to food security, sustainable land use and climate resilience.

The Leaf Area Index (LAI) is one of the most important biophysical parameters used in smart agriculture for assessing plant growth and nutrient status. LAI characterises the canopy structure of crops and is intricately linked to vital plant processes such as photosynthesis, respiration and transpiration [4]. Accurate estimation of LAI is therefore essential for understanding plant health, productivity and ecosystem functioning.

In remote sensing applications, biophysical parameters such as LAI are often estimated using vegetation indices. These indices are derived from specific combinations of spectral bands that correlate with vegetation characteristics. Among them, the most commonly used are Normalised Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI) and others (described below). Although these indices do not measure biophysical quantities like LAI directly, they are extensively employed due to their empirical correlations with structural and physiological characteristics of vegetation, including biomass, chlorophyll content and canopy density.

Traditionally, LAI has been measured through field sampling and manual techniques, which, while accurate, are time-consuming, labor-intensive and unsuitable for large-scale or frequent monitoring. The increasing availability of high-resolution remote sensing data from satellites and unmanned aerial vehicles (UAVs) has enabled more scalable and efficient approaches to LAI estimation [105, 115]. The use of multispectral and hyperspectral imagery, combined with advanced machine learning techniques, has shown great potential to improve the accuracy and reliability of LAI prediction, overcoming the limitations of traditional vegetation indices such as NDVI [42].

Recent studies have further explored the integration of machine learning models with high-resolution remote sensing data to estimate vegetation-related biophysical parameters more accurately [41, 113]. These approaches often demonstrate enhanced predictive performance compared to traditional empirical methods, particularly when applied to specific vegetation types or ecosystems, where machine learning regressions like Gaussian Process Regression or Extra Trees demonstrate superior predictive performance compared to linear models [33, 63]. Building on these developments, the present study investigates the implementation of a neural network model for LAI estimation using high-resolution multispectral imagery acquired from different satellites. Such approaches contribute to the broader effort of developing reliable, scalable, and non-destructive methods for vegetation monitoring and precision agriculture applications.

Overall, the utilisation of multispectral technology with high spatial resolution, in tandem with advanced machine learning models, offers a promising pathway for rapid, non-destructive, and accurate monitoring of LAI and other biophysical properties [10, 58, 108]. This advancement not only aids in precision agriculture [94] but also contributes significantly to the broader field of remote sensing and plant health assessment. Studies [3, 73] conclude that machine learning algorithms represent a robust and efficient approach for predicting biodiversity-related variables from remote sensing data. In light of this, the present work explores the implementation of a neural network model capable of estimating LAI based on high-resolution multispectral imagery acquired from the multiple satellite platforms.

2.2 Satellite Imagery and its Usage for Earth Observation

This section provides an overview of types of satellite data commonly used for Earth observation and their potential for land cover monitoring. The characteristics of the data collected by some of the satellites that are the focus of the study are described. An overview of the vegetation indices is made including their relationship with the corresponding biophysical properties.

2.2.1 Satellite Data

Remote sensing satellite missions are essential for land conservation and global sustainability by providing crucial data through multispectral imaging. Such images captures spectral signatures that represent distinct land cover features across the electromagnetic spectrum [84]. The increase in worldwide open data satellite missions has significantly expanded the use of remote land observation, leading to widespread accessibility to satellite imagery [2].

This background review is intentionally limited to spectral regions, sensor types, and resolution properties that directly influence vegetation monitoring and machine learning-based biophysical variable estimation.

According to the wavelength ranges of the electromagnetic waves, satellite remote sensing data can be grouped into several types (Figure 2.3). In the context of Earth surface observation and vegetation monitoring, optical and infrared wavelengths are of primary importance, while other regions of the spectrum are typically used for space science or atmospheric studies and are therefore outside the scope of this work.

- **Ultraviolet (UV) data:** Ultraviolet radiation occupies the spectral range between X-rays and visible light (100–400 nm) and is commonly divided into UVC, UVB and UVA regions. While UV radiation affects plant physiology, UVC and most UVB radiation are absorbed by atmospheric gases and do not reach the Earth's surface under normal conditions [11]. Consequently,

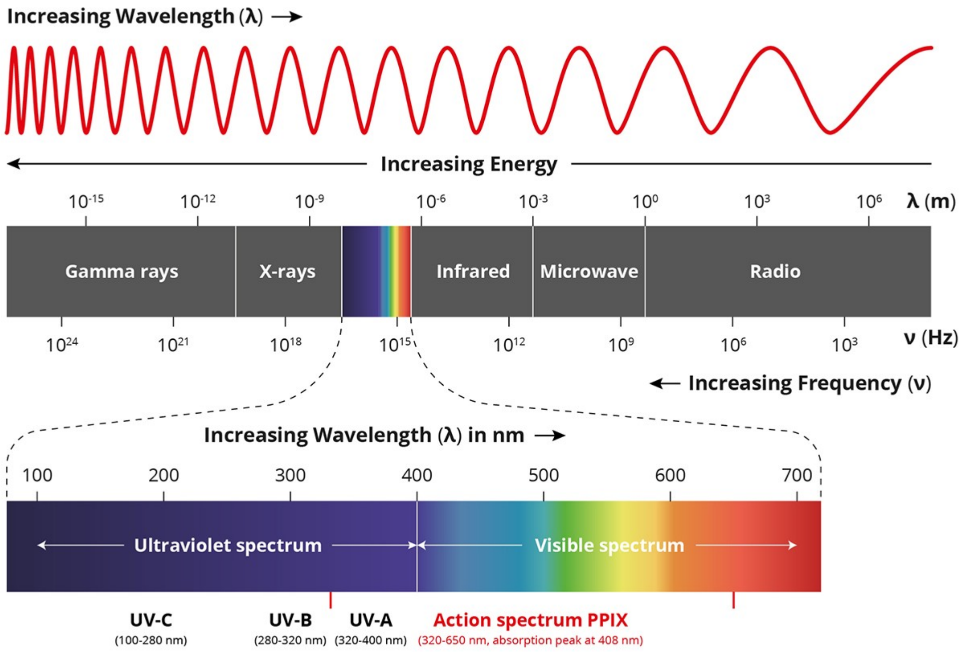


FIGURE 2.3: Electromagnetic spectrum
(source: Secrets of the Universe).

UV-based satellite observations are rarely used for large-scale vegetation monitoring and are not considered further in this study.

- **Optical data:** Optical data include wavelengths in the visible spectrum, approximately from 400 nm to 700 nm. These wavelengths are fundamental for visual interpretation and form the basis for many vegetation indices and land cover classification approaches.
- **Infrared data:** Infrared radiation covers wavelengths located between visible light and microwaves and is typically divided into several regions. Among these, the most relevant for vegetation monitoring are the near-infrared (NIR) and the shortwave infrared (SWIR) regions. The NIR region generally includes wavelengths from approximately 700 to 1400 nm, while the SWIR region covers wavelengths from about 1400 to 3000 nm. NIR wavelengths are particularly sensitive to vegetation structure and biomass, while SWIR bands provide information related to moisture content, vegetation stress, and soil properties. These spectral regions play a central role in vegetation index computation and biophysical variable estimation. Mid-wave and long-wave infrared wavelengths are primarily used for thermal applications and are therefore outside the scope of this work.
- **Radar data:** Radar systems operate in the microwave region (1 mm to 1 m) and enable all-weather observations due to their ability to penetrate clouds and atmospheric disturbances. Although radar data are valuable for topographic mapping and soil moisture estimation, this thesis focuses on passive optical multispectral data, and radar-based observations are not considered further.

In addition to spectral imagery, other sensors such as LiDAR, satellite altimetry, and GNSS systems provide complementary information for terrain modelling, geolocation, and Earth system monitoring. These data are outside the scope of this work, which concentrates on pixel-based analysis of optical satellite imagery.

Based on the spectral characteristics described above, satellite imagery can be categorized into different types, each offering distinct advantages and limitations for vegetation monitoring and land cover analysis [125]:

- **Panchromatic images (PAN):** Panchromatic images record data in a broad spectral band covering the visible and near-infrared regions. They are single-channel grayscale images with high spatial resolution and are mainly used for visual interpretation and spatial detail enhancement.
- **Multispectral images (MS):** Multispectral images record data in multiple discrete spectral bands, typically including visible and near-infrared wavelengths and sometimes shortwave infrared bands. Different materials exhibit distinctive spectral signatures, which makes multispectral data particularly suitable for vegetation monitoring, land cover classification and environmental analysis.
- **Hyperspectral images (HS):** Hyperspectral images capture information in a large number of contiguous spectral bands, often spanning the visible, NIR and SWIR regions. Each pixel contains a detailed spectral signature, typically represented as a three-dimensional data cube with two spatial dimensions and one spectral dimension. Although hyperspectral data offer rich spectral information, their large data volume, complex processing requirements and limited availability restrict their widespread use in large-scale operational applications.

Beyond the type of imagery, the effectiveness of satellite data also depends on its spatial, spectral, and radiometric resolution, which collectively determine the level of detail and accuracy achievable in vegetation assessments.

The spatial resolution of a remote sensing imaging system is defined by the area of the ground represented by a single pixel and directly affects the level of detail captured within a scene. As pixel size decreases, finer spatial details are preserved in the digital representation [66].

Nadir-viewing geometry further influences spatial resolution by minimising geometric distortions and variations in pixel footprint across the scene [86].

Spectral resolution describes the bandwidth of the electromagnetic signals captured by a sensor. Narrower spectral bands correspond to higher spectral resolution. Multispectral sensors typically acquire data in a limited number of bands, while hyperspectral sensors may contain hundreds of narrow spectral channels [74].

Radiometric resolution refers to the sensitivity of a sensor to detect small differences in energy levels and is commonly expressed in bits. Higher radiometric resolution allows finer discrimination between surfaces with similar spectral characteristics.

In summary, panchromatic images provide high spatial resolution but contain limited spectral information, whereas multispectral images offer richer spectral content at lower spatial resolution. Freely available satellite imagery generally exhibits coarser spatial resolution compared to commercial platforms, which can introduce challenges for detailed spatial analysis. Hyperspectral imagery combines high spectral richness with good spatial resolution but requires substantial storage capacity and computational resources, as well as specialised sensors and platforms.

2.2.2 Satellites and Sensors

During the last few decades, the availability of high spatial, temporal, spectral and radiometric resolution imagery from low-orbiting small satellites has greatly improved, widening the remote sensing applications further with better precision and accuracy. Especially, satellites such as GeoEye, IKONOS, KOMPSAT-3, Landsat-8, Pleiades 1A, Pleiades 1B, QuickBird, Sentinel-2 and WorldView-3 have been effectively utilised in many agricultural applications [124]. These applications encompass but are not restricted to the retrieval of vegetation indices [99], the estimation of plant height [122], evapotranspiration [85], leaf/canopy chlorophyll content [45], yield [64], and crop-health monitoring [25].

The diversity in spatial and temporal resolutions, spectral characteristics (Table 2.1) and the varying spatial shapes of output images produced by different satellite sensors complicate the application of deep learning algorithms to remote sensing imagery. These variations require the development of versatile and adaptable models capable of handling datasets with heterogeneous resolutions, dimensionalities and data quality.

TABLE 2.1: Spatial and temporal resolution and spectral differences between different satellite probes.

Satellite	Spatial resolution PAN (m)	Spatial resolution MS (m)	Spectral bands (#)	Temporal resolution (days)	Radiometric resolution (bits)	Swath width (km)	Wavelength range (nm)
GeoEye-1	0.41	1.65	4	1–3	11	15.2	450–800
IKONOS	0.82	3.2	4	1–3	11	11.3	445–853
KOMPSAT-3	0.7	2.8	5	1	14	15	450–900
Landsat 8	15	30	11	16	16	185	433–12500
Pleiades 1A, 1B	0.5	2	5	1–2	12	20	430–950
QuickBird	0.65	2.62	4	1–3.5	11	16.5	450–900
Sentinel-2	10	20	13	5	12	290	442–2186
WorldView-2	0.46	1.84	8	1.1	11	16.4	400–1040
WorldView-3	0.31	1.24	28	1	11	13.1	397–2373

(source: [84, 125])

In addition to these structural differences, the relationships between spectral observations and biophysical variables are inherently non-linear, driven by complex interactions between vegetation properties, illumination conditions, atmospheric effects and sensor characteristics. Such non-linearities further challenge traditional modelling approaches and motivate the use of deep learning methods, which are well suited to learning complex, non-linear mappings directly from high-dimensional remote sensing data.

Although the framework is not limited to any specific sensor pair, this work focuses on Sentinel-2 and Landsat 8/9 as the two primary freely available medium-resolution satellite missions with global coverage and operational continuity.

Sentinel-2

Sentinel-2 [21], a part of the Copernicus Program, is equipped with multi-spectral imaging across 13 spectral bands with varying spatial resolutions, effectively capturing a broad spectrum of Earth's features. The bands are divided into four categories based on their resolution: four bands at 10 meters (including visible and near-infrared), six bands at 20 meters (covering red-edge, shortwave infrared and additional near-infrared) and three bands at 60 meters (including a band for atmospheric applications). The 10-meter resolution bands are particularly noteworthy as they offer a detailed view of the Earth's surface, suitable for monitoring urban areas and agricultural lands. Meanwhile, the 20 and 60-meter bands provide a helpful base for analysing vegetation and water bodies, as well as aiding in atmospheric correction. This diverse range of bands, along with their respective resolutions and pixel sizes, makes Sentinel-2 highly versatile for a variety of applications, ranging from land monitoring and disaster management to agricultural mapping and climate change studies.

The Sentinel-2 constellation consists of two identical satellites, Sentinel-2A launched in 2015 and Sentinel-2B launched in 2017, which together achieve a global revisit frequency of five days.

Landsat 8/9

Landsat 8 [78] entered service in 2013 with enhanced thermal and infrared capabilities, marking a new era in Earth observation in terms of providing free and open data for monitoring the Earth's landscape. Landsat 8 operates at an altitude of 705 km in a sun-synchronous orbit. It orbits the Earth every 99 minutes. With an inclination angle of 98.2 degrees, it covers every point on Earth globally every 16 days. Its sensors OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) operate with 11-band spectral resolution and 12-bit radiometric resolution. The spatial resolution is 30 m (VNIR and SWIR) and 15 m (panchromatic), with scenes measuring 185 x 180 km and generating a total of over 700 scenes per day.

Landsat 9 [79], launched in September 2021, shares most of the orbital and spatial characteristics of Landsat 8 but carries improved OLI-2 and TIRS-2 sensors with enhanced radiometric resolution . When both satellites operate together, the effective global revisit time is reduced from 16 to 8 days.

As shown in Figure 2.4, the spectral bands of Sentinel-2 and Landsat 8/9 differ significantly in their wavelength coverage.

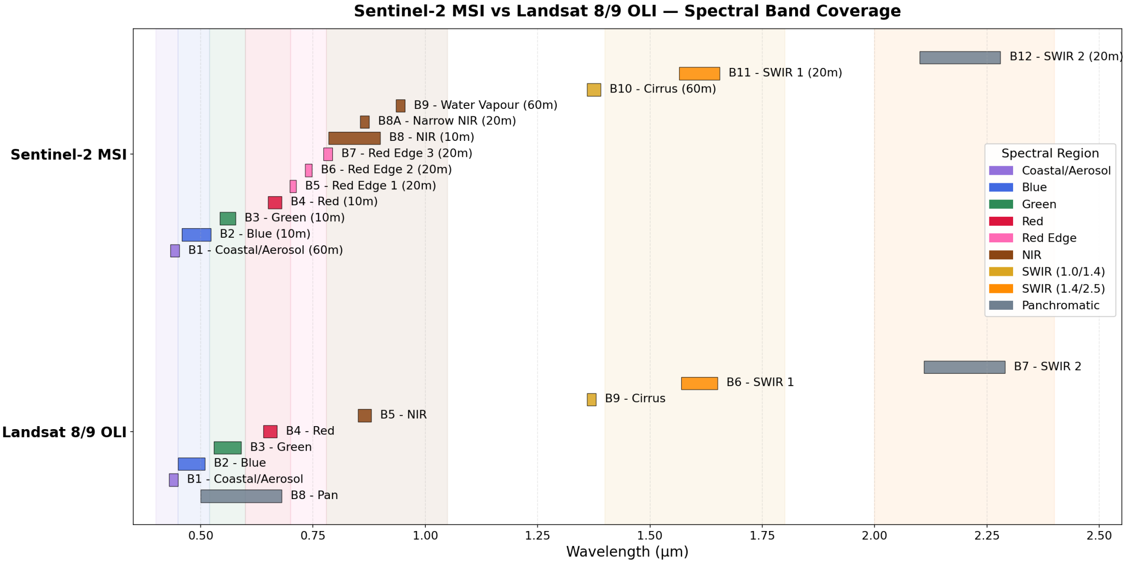


FIGURE 2.4: Spectral response comparison Sentinel-2 and Landsat 8/9.

2.2.3 Biophysical Properties

The term **Biophysical properties** refers to the physical characteristics and behaviours of biological materials or systems that can be measured using different physics techniques. These properties encompass a wide array of physical and chemical characteristics that are inherent to biological systems, such as mechanical strength, electrical conductivity, thermal properties and molecular dynamics. By studying these properties, valuable insight into the functioning and behaviour of living organisms is obtained.

Leaf Area Index

Leaf Area Index (LAI) is a fundamental biophysical parameter defined as the total one-sided leaf area per unit of horizontal ground area [16]:

$$\mathrm{LAI} = \frac{LeafArea}{GroundArea}$$

where $GroundArea$ means “horizontal land surface area” to clarify LAI on a sloping surface and the definition of $LeafArea$ depends on the type of plants due to differences in their leaf characteristics: in broadleaf canopies it is defined as one-sided green leaf area, while in canopies three definitions are used – half of the total needle surface area, one-sided needle area (similar to the broadleaf canopies), and total needle surface area.

It is a dimensionless value, but can also be expressed in units of measurement such as cm^2/cm^2, m^2/m^2, in^2/in^2, etc. The value is transcribed in the following manner: LAI = 1.5 means there are 1.5 square meters of leaf area per square meter of ground area. Globally LAI is highly variable – in some desert areas the value of LAI is close to 0, while in the densest tropical forests it can be more than 9. Mid-latitude forests and scrub typically have LAI values between 3 and 6. Within a season, cultivated areas also show large variations in LAI. For example, from seeding to maturity, the leaf area index of maize ranges from 0 to 6 at maturity. For these reasons, LAI appears to be a useful metric for describing spatial and temporal patterns of canopy growth and productivity.

LAI quantifies canopy density, leaf layering and total foliage biomass, and is therefore directly linked to processes such as photosynthesis, transpiration and radiation interception. Because of its physical interpretation and its strong relevance for ecological and agricultural modelling, LAI is widely used as a structural indicator of vegetation condition [5, 75, 124].

Direct measurements of LAI involve destructive sampling, litterfall collection, or point contact methods. Indirect methods involve ground-based, lidar or remote sensing instruments (presented in detail in [17]).

Spectral vegetation indices (some of which we will present below) can provide indirect estimates of LAI by exploiting the contrast between visible and near-infrared reflectance. However, these indices tend to saturate at high canopy densities [6, 38].

Among model-based methods, radiative transfer models such as PROSPECT [53] and SAIL [106] simulate leaf and canopy optical properties to retrieve LAI from reflectance measurements. These physically-based models provide a more accurate link between spectral observations and canopy structure than simple indices, particularly in dense or heterogeneous canopies. Supervised regression methods calibrated against field measurements can also be used to map LAI from multispectral imagery.

Biomass

Biomass measures the total living plant material including trees, shrubs and ground vegetation. Biomass can be quantified via direct field measurements (e.g., harvest method), mathematical models using easily measurable parameters such as tree height and crown width, or remote sensing techniques based on spectral vegetation indices. Hybrid approaches often combine field and remote sensing data to calibrate and validate biomass estimates, making them particularly suitable for large-scale agricultural or forestry applications [72].

Chlorophyll Content

Chlorophyll content is a key indicator of plant health and photosynthetic activity. It is widely used to detect stress conditions, assess nutrient status, and monitor the impact of environmental changes. Chlorophyll can be estimated using spectral indices such as the Chlorophyll Index (CI) or the Canopy Chlorophyll Content Index (CCCI), which leverage green and red-edge bands to derive chlorophyll concentrations remotely [12].

Soil Moisture Content

Soil moisture content represents the amount of water in the soil and is essential for plant growth, affecting water availability to roots and soil chemical and biological processes. Soil moisture content can be expressed as a volume percentage (Volumetric Water Content – measures the volume of water in the soil relative to the total volume of the soil) or as a weight percentage (Gravimetric Water Content – measures the mass of water in the soil relative to the dry mass of the soil) or even as soil water potential (Soil Water Potential – measures the energy required to remove water from the soil). Soil moisture content measurement methods include direct and indirect in-situ methods, as well as remote sensing using drones for high-precision local data and satellites with radiometric or microwave sensors to estimate the content of soil moisture over large areas [30].

Leaf Water Content

Leaf water content (LWC) describes the amount of liquid water contained within plant leaves and is an important physiological parameter related to plant hydration, transpiration and photosynthetic activity. Variations in leaf water content are closely associated with environmental conditions and plant stress, particularly drought and insufficient irrigation [15].

Leaf water content can be expressed through measures such as relative water content (RWC) or equivalent water thickness (EWT), which represent the amount of water per unit leaf area. Direct measurements typically require destructive sampling and laboratory analysis. However, remote sensing techniques enable indirect estimation of vegetation water status over large areas. Spectral indices that use near-infrared (NIR) and shortwave-infrared (SWIR) wavelengths, such as the Normalized Difference Water Index (NDWI), exploit the strong absorption of water in the SWIR region to infer vegetation water content.

Biophysical properties provide comprehensive insights into the structure, function, and health of biological systems. They allow researchers and practitioners to monitor ecosystems, guide agricultural management, and assess environmental change. Advances in remote sensing and modeling techniques have greatly improved the ability to measure these properties, enabling large-scale monitoring and timely decision-making.

2.2.4 Vegetation Indices

Many biophysical properties, that are used by researchers in monitoring plant cover qualities – such as percentage of vegetation cover, green leaf biomass or leaf area index (LAI) – can be presented by spectral **Vegetation Indices (VIs)**. VIs are generated by combining data from multiple spectral bands into a single value [76]. The data can be collected using handheld sensors or remote sensing instruments mounted on airborne vehicles (aircraft or drones) or satellites. Such instruments measure the reflectance of different wavelengths of light from the Earth's surface. The VIs typically exploit specific spectral bands to highlight the contrast between vegetation and other surfaces [36]. Chlorophyll spectral response, shown in Figure 2.5, illustrates how different wavelengths are absorbed and reflected by vegetation.

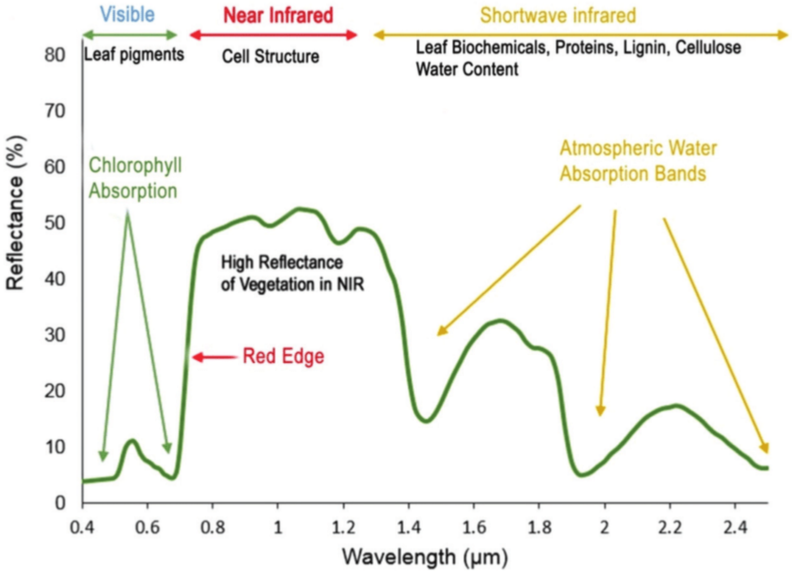


FIGURE 2.5: Chlorophyll spectral response [88].

Some VIs use reflectance values only at visible wavelengths (380 nm to 700 nm), such as Excess Greenness Index (EGI), Green Chromatic Coordinate (GCC), Triangular Greenness Index (TGI) and so on. RGB-based indices are mainly useful for estimating overall green biomass or canopy cover, but they provide limited information on leaf physiology or structure.

For assessing vegetation biophysical and biochemical traits, in particular LAI, broad-spectrum spectral indices that incorporate NIR and red-edge bands are commonly used to capture canopy structure and physiological status more accurately.

Three of the most commonly used spectral vegetation indices, selected for this study due to their effectiveness in capturing key vegetation reflectance traits and their suitability for analysis with Landsat and Sentinel sensors, are the Normalized Difference Vegetation Index (**NDVI**), the Green Normalized Difference Vegetation Index (**GNDVI**), and the Soil-Adjusted Vegetation Index (**SAVI**).

Normalised Difference Vegetation Index

One of the most widely used indices for assessing vegetation activity, developed in the early days of remote sensing research, was introduced by Rouse and colleagues in 1973 [90] and later became known as the NDVI.

NDVI is calculated as the normalised difference between reflected Near-infrared (ρ_{NIR}) and reflected Red light (ρ_{Red}):

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{Red}}}{\rho_{\text{NIR}} + \rho_{\text{Red}}} \tag{2.1}$$

Characteristics:

- Sensitive to chlorophyll absorption and leaf scattering;
- Known saturation at LAI $\approx$ 3–4 m^2/m^2 for dense canopies;
- Validation performance: $R = 0.79$, RMSE $= 0.77$ m^2/m^2 (best overall).

NDVI values range from -1 to +1. Negative values usually mark water areas. Low positive values (0.1 and below) correspond to barren areas of rock, sand or urban areas. Moderate values represent shrub and grassland (0.2 to 0.3). High values indicate temperate and tropical rainforests (0.6 to 0.8) [76].

NDVI is sensitive to atmospheric conditions and therefore needs atmospheric corrections for its accurate calculation. Other weaknesses is the non-linear relationship of NDVI with the biophysical parameters of vegetation, such as LAI - at high and low values of these parameters, NDVI loses its sensitivity. It also provides a general measure of greenness, but cannot provide more detailed information about specific plant species or their age and health status.

Despite these weaknesses, NDVI is one of the most commonly used indices in remote sensing of the Earth's surface and in particular in vegetation monitoring due to its easy calculation and its sensitivity to vegetation cover and plant health. The NDVI is a standard index used by many satellite missions and there is already a huge amount of historical data well covering the Earth's surface that allows comparisons and analyses of long-term trends in vegetation.

Green Normalised Difference Vegetation Index

For improving the sensitivity to the chlorophyll content in the plant biomass Gitelson and Merzlyak in 1996 [39] suggest the modification of NDVI using Green instead of Red channel:

$$\mathrm{GNDVI} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{Green}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{Green}}} \tag{2.2}$$

Characteristics:

- Substitutes green band for red, more sensitive to chlorophyll concentration;
- Later saturation point (LAI $\approx$ 4–5 m^2/m^2) due to deeper canopy penetration;
- Validation performance: $R = 0.73$, RMSE $= 0.82$ m^2/m^2.

GNDVI also takes values from -1 to 1 which are interpreted similarly to those of NDVI. The use of green spectral band in GNDVI makes it more sensitive to chlorophyll content in plants compared to NDVI, which is particularly useful for applications related to plant health and condition [96].

Soil-Adjusted Vegetation Index

SAVI was developed by Huerte in 1988 [49] to more accurately model NIR radiant transfer in incomplete vegetation canopies to remove canopy background noise:

$$\text{SAVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{Red}}}{\rho_{\text{NIR}} + \rho_{\text{Red}} + L} \times (1 + L) \tag{2.3}$$

where L is the soil brightness correction factor, typically set to 0.5 for intermediate vegetation density.

Characteristics:

- Minimizes soil brightness influence in partially vegetated pixels;
- Particularly effective for sparse canopies and early growing season;
- Validation performance: $R = 0.77$, $\text{RMSE} = 0.85\ \text{m}^2/\text{m}^2$.

SAVI values range from -1 to +1, where negative values correspond to non-vegetated surfaces, while near 1 values indicate extremely dense vegetation. SAVI is particularly effective in areas with sparse vegetation cover since the influence of soil reflection is minimised through the correction factor L. The formula is easy to apply and uses standard data available in all standard multispectral imaging sensors. Despite the correction of the influence of the soil, the single correction factor can lead to inaccuracies due to the lack of adaptability to different types of soils and conditions.

2.3 Correction and Enhancing

2.3.1 Athmospheric Correction

Satellite observations are influenced by interactions between solar radiation and atmospheric constituents, including scattering, absorption, and emission processes. As a result, the signal recorded by satellite sensors represents the radiance at the Top of Atmosphere (TOA), which includes contributions from both the Earth's surface and the atmosphere. In remote sensing literature, three reflectance representations are commonly distinguished: Top-of-Atmosphere (TOA) reflectance, Bottom-of-Atmosphere (BOA) or surface reflectance, and Bidirectional Reflectance Distribution Function (BRDF)-normalized reflectance. TOA reflectance represents the signal recorded by the sensor and includes atmospheric contributions. BOA reflectance attempts to isolate the surface signal after atmospheric correction. BRDF-normalized reflectance further accounts for directional reflectance effects caused by sun–sensor geometry. To retrieve the actual surface reflectance, commonly referred to as Bottom of Atmosphere (BOA) reflectance, atmospheric correction procedures are applied to remove atmospheric effects and transform TOA measurements into physically meaningful surface reflectance products (see Figure 2.6).

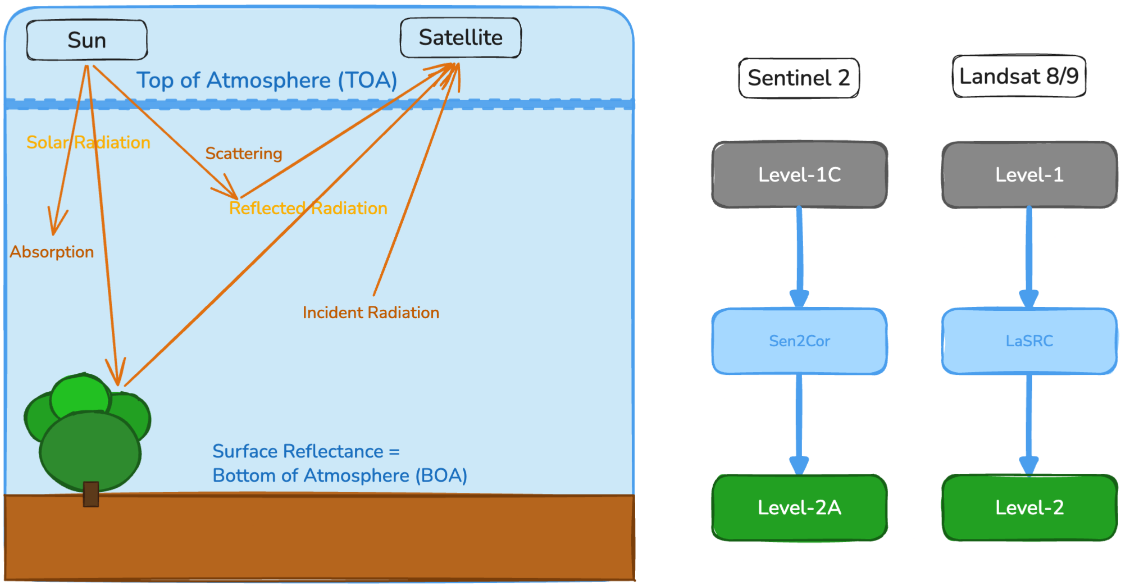


FIGURE 2.6: Conceptual diagram of atmospheric effects on satellite observations, illustrating the transformation from TOA measurements to BOA/surface reflectance through atmospheric correction (Level-1/Level-2 processing).

Different satellite sensors generate BOA/surface reflectance products using sensor-specific atmospheric correction algorithms. This section compares the approaches used for Sentinel-2 and Landsat 8, highlighting their suitability for LAI retrieval and subsequent vegetation analysis.

Sentinel 2

Sentinel-2 data products are distributed at several processing levels, with the distinction between Level-1 and Level-2 being particularly important for quantitative terrestrial applications such as Leaf Area Index (LAI) estimation. Level-1 products provide TOA reflectance, while Level-2 products provide BOA (surface) reflectance, obtained through atmospheric correction. The operational Sentinel-2 Level-2 product is generated from Level-1 using the Sen2Cor processor, which is developed and maintained within the Copernicus ground segment framework [32, 70].

Level-1 products: TOA reflectance and geometric correction

Sentinel-2 Level-1 products provide orthorectified TOA reflectance for all multispectral bands, delivered in a cartographic geometry based on a global digital elevation model [32]. At this stage, radiometric processing accounts for sensor calibration, stray light correction and spectral response normalization, but atmospheric effects such as molecular scattering, aerosol scattering and gaseous absorption are not removed. Consequently, Level-1 reflectance values still contain contributions from the atmosphere, which can introduce variability unrelated to surface properties, particularly in the visible and near-infrared bands that are critical for vegetation analysis [70].

For LAI estimation, the use of TOA reflectance is generally suboptimal unless atmospheric effects are explicitly modeled within the retrieval approach. Atmospheric variability can propagate into vegetation indices and physically based inversion schemes, leading to biases and reduced temporal consistency.

Level-2 products: surface reflectance and auxiliary layers

The Sentinel-2 Level-2 product provides atmospherically corrected surface reflectance derived from Level-1 data. According to the Sentinel-2 Level-2 Product Definition Document [32], the Level-2 processing chain performs atmospheric correction, scene classification, and retrieval of auxiliary atmospheric parameters, generating outputs that include surface reflectance for all spectral bands, a Scene Classification Layer (SCL) identifying clouds, cloud shadows, snow/ice, water, vegetation, and non-vegetated surfaces, and spatially distributed estimates of Aerosol Optical Thickness and total column Water Vapour. These additional layers are intended to support quality control and uncertainty assessment in downstream applications, including biophysical variable retrieval.

Algorithm design objectives and limitations

Sen2Cor is a physically based atmospheric correction algorithm that relies on radiative transfer modeling and precomputed look-up tables (LUTs). The LUTs are generated using the *libRadtran* radiative transfer model and DISORT solver and are parameterised across atmospheric profiles, aerosol models, solar–view geometries and surface elevation [31]. During processing, TOA reflectance values are inverted using these LUTs to estimate surface reflectance by accounting for Rayleigh scattering, aerosol scattering & absorption and gaseous absorption by ozone, oxygen and water vapour.

The Sentinel-2 Level-2 atmospheric correction was optimised to provide globally consistent, operational surface reflectance products suitable for a wide range of land applications, including vegetation monitoring, land cover mapping and environmental change detection [32, 70]. The algorithm emphasises physical consistency, scalability and the availability of auxiliary quality layers, rather than optimisation for any single downstream variable such as LAI.

A key limitation for time-series vegetation studies is the absence of Bidirectional Reflectance Distribution Function (BRDF) correction. Directional reflectance effects related to changing sun–sensor geometry can introduce systematic variability in red, near-infrared and red-edge reflectance, which may propagate into LAI estimates. Additionally, published evaluations indicate that cloud and shadow detection performance of Sen2Cor can vary with surface type and atmospheric conditions, potentially affecting the reliability of biophysical retrievals if masking is not carefully refined [97].

Landsat 8

Landsat 8 data products are distributed at multiple processing levels, with Level-1 products providing TOA reflectance and Level-2 products providing atmospherically corrected BOA reflectance. As for Sentinel-2, this distinction is critical for quantitative vegetation analysis, including LAI estimation. The operational Landsat 8 Level-2 Surface Reflectance product is generated from Level-1 inputs using the Land Surface Reflectance Code (LaSRC), which constitutes the standard atmospheric correction framework for Landsat Operational Land Imager (OLI) data [103, 107].

Level-1 products: TOA reflectance

Landsat 8 Level-1 products consist of radiometrically calibrated and geometrically corrected TOA reflectance values, corrected for sensor-specific effects and precise geolocation. As with Sentinel-2 Level-1 products, atmospheric effects are not removed at this stage and TOA reflectance therefore retains contributions from molecular scattering, aerosols and gaseous absorption [107]. These effects can introduce variability unrelated to vegetation state, particularly in the visible and near-infrared spectral regions relevant for LAI retrieval.

Level-2 products: surface reflectance and quality information

Landsat 8 Level-2 Surface Reflectance products provide atmospherically corrected BOA reflectance derived from Level-1 TOA inputs. According to the Landsat Collection 2 product documentation, the Level-2 processing chain includes atmospheric correction and the generation of quality assessment layers, such as cloud, cloud-shadow, snow and water masks, which support data screening and temporal consistency analyses [103].

In contrast to Sentinel-2 Level-2 products, Landsat Level-2 products do not routinely provide spatially explicit user-facing layers for Aerosol Optical Thickness or Water Vapour. Instead, atmospheric parameters are estimated internally within the correction process and are reflected implicitly in the corrected surface reflectance and associated quality flags.

Algorithm design objectives and limitations

LaSRC follows the same general physically based radiative transfer principles described for Sentinel-2 Level-2 atmospheric correction (Section 2.3.1), using the 6S radiative transfer model (Second Simulation of a Satellite Signal in the Solar Spectrum) to correct TOA reflectance for molecular scattering, aerosol effects and gaseous absorption [107]. The key design objective of LaSRC is the production of consistent and stable surface reflectance records across multiple Landsat missions, enabling long-term land surface and vegetation monitoring.

Accordingly, the Landsat Level-2 atmospheric correction emphasises robustness and cross-mission temporal continuity rather than optimisation for a specific downstream biophysical variable such as LAI [103, 107]. As with Sentinel-2 Level-2 products, Landsat surface reflectance is directional and does not include bidirectional reflectance distribution function (BRDF) correction, meaning that variations in sun–sensor geometry can introduce reflectance variability unrelated to true surface or vegetation changes.

Synthesis and Methodological Decision

Across Sentinel-2 and Landsat 8, the atmospheric correction strategies adopted in this study reflect a balance between physical rigor, operational robustness and empirical suitability for Leaf Area Index (LAI) estimation. Despite differences in implementation and product maturity, all sensors rely on physically based radiative transfer principles to convert Top-of-Atmosphere (TOA) reflectance to Bottom-of-Atmosphere (BOA) or surface reflectance, which is a necessary precondition for reducing atmospheric-induced variability in vegetation-sensitive spectral bands.

Empirically, numerous validation studies have demonstrated that surface reflectance products derived from operational processors such as Sen2Cor (Sentinel-2) and LaSRC (Landsat 8) provide substantially improved temporal stability and physical interpretability for vegetation indices and biophysical variable retrieval compared to TOA reflectance, particularly in the red, near-infrared and red-edge spectral regions that dominate LAI sensitivity [70, 107]. These improvements are especially relevant for multi-temporal and multi-sensor analyses, where uncorrected atmospheric variability would otherwise confound vegetation dynamics.

At the same time, it is recognised that none of the operational surface reflectance products considered here are explicitly optimised for LAI estimation, nor do they include bidirectional reflectance distribution function (BRDF) correction. As documented for both Sentinel-2 and Landsat Level-2 products, surface reflectance remains directional and variations in illumination and viewing geometry can introduce systematic effects unrelated to canopy structure [32, 107]. These limitations are well known in the literature and are commonly addressed through temporal compositing, quality filtering or downstream modelling choices rather than through modifications to the atmospheric correction itself.

Taken together, the adopted atmospheric correction approaches provide a consistent and empirically supported foundation for LAI estimation across sensors, while acknowledging known limitations related to BRDF effects and sensor-specific processing differences. Given the broader objectives of this thesis – namely, the development and evaluation of sensor-agnostic representations for LAI estimation that generalize across regions and time periods – the use of established operational surface reflectance products is methodologically justified. Further investigation into alternative atmospheric correction or reflectance normalization strategies, while valuable, would constitute an extension beyond the defined scope of this work and is therefore not pursued here [70, 107].

2.3.2 Illumination and Viewing Geometry

The reflectance measured by optical satellite sensors is inherently dependent on illumination and viewing geometry, in addition to surface properties. Vegetation canopies are anisotropic reflectors, meaning that the recorded signal varies with solar zenith angle, sensor viewing zenith angle and the relative azimuth between illumination and observation directions. These angular dependencies arise from canopy structure, leaf orientation, shadowing and multiple scattering processes and can introduce variability in surface reflectance that is not directly related to changes in biophysical variables such as Leaf Area Index (LAI) [27, 92].

The Figure 2.7 illustrates the dependence of reflectance on illumination and viewing angles.

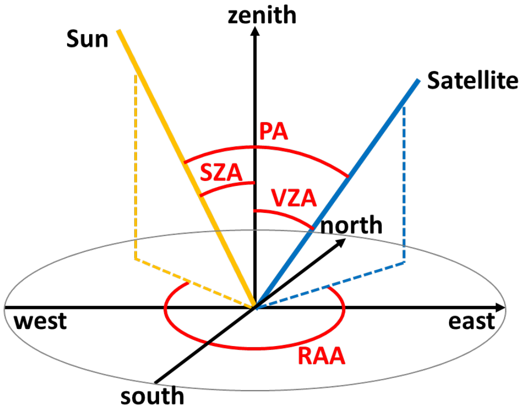


FIGURE 2.7: Geometry of illumination and viewing angles in remote sensing.

The *solar zenith angle* (SZA), defined as the angle between the sun and the local vertical, controls the illumination conditions of the canopy. Larger SZAs, corresponding to lower sun elevations, increase shadow length and modify the balance between sunlit and shaded foliage. This can affect reflectance in the visible and near-infrared spectral regions, which are critical for LAI-sensitive vegetation indices and inversion-based retrieval methods [77, 92]. Seasonal and latitudinal variations in SZA therefore represent a primary source of angular variability in multi-temporal vegetation observations.

The *sensor viewing zenith angle* (VZA), defined as the angle between the sensor line of sight and the local vertical, determines which portions of the canopy are observed. Near-nadir observations tend to minimise anisotropic effects, whereas off-nadir views enhance directional scattering and shadowing, increasing reflectance variability unrelated to vegetation state. The influence of VZA is particularly pronounced in structurally complex canopies and at high spatial resolution, where sub-pixel heterogeneity and adjacency effects become more important [27].

The *relative azimuth angle* (RAA), describing the angular difference between solar and sensor azimuth directions, further modulates reflectance anisotropy by controlling forward and backward scattering regimes. Certain combinations of viewing and illumination geometry can reduce apparent shadowing (hot-spot effects) or accentuate it, thereby influencing observed reflectance independently of LAI [92].

The importance of these angular effects varies across sensors. Sentinel-2 and Landsat 8 acquire imagery with near-nadir viewing geometries and relatively limited variation in VZA, meaning that angular variability is dominated by changes in solar illumination rather than viewing direction. In contrast, WorldView-3 acquisitions often include substantial off-nadir viewing angles, making viewing geometry a first-order contributor to reflectance variability. This difference is particularly relevant for high-resolution data, where local shadowing and canopy heterogeneity amplify angular effects [27]. For all three sensors, the surface reflectance products considered in this study are directional and do not include explicit BRDF correction, implying that residual angular effects remain in the input data.

Implications for LAI Estimation and Robustness

Despite the presence of angular effects in surface reflectance, LAI is generally considered more robust to illumination and viewing geometry than instantaneous spectral indices. LAI represents a structural canopy property that varies relatively slowly in time compared to reflectance and retrieval approaches typically exploit multispectral information and non-linear relationships that reduce sensitivity to single-band directional effects [77, 118]. Empirical studies have shown that, while angular variability can introduce noise and bias in reflectance, its impact on LAI retrieval is often secondary compared to uncertainties arising from atmospheric effects, cloud contamination or model assumptions, particularly when quality filtering and temporal aggregation are applied [27, 118]. Consequently, while angular effects are acknowledged as a source of residual variability in this study, LAI is treated as sufficiently robust for the comparative and multi-sensor analyses undertaken here, without explicit BRDF normalization.

2.3.3 Pansharpening

Despite continuous improvements in multispectral satellite sensors, a fundamental trade-off remains between spatial and spectral resolution. High-resolution panchromatic (PAN) bands are often acquired alongside lower-resolution multispectral (MS) bands, motivating the use of pansharpening techniques to enhance the spatial resolution of multispectral imagery while attempting to preserve its spectral fidelity. This process is particularly relevant in applications requiring detailed spatial representation, but it also introduces potential risks when quantitative biophysical variables, such as Leaf Area Index, are derived from the fused data.

Classes of Pansharpening Methods

Pansharpening approaches can be broadly categorised into classical signal-processing methods and learning-based methods. Classical techniques include component substitution methods, such as Intensity-Hue-Saturation (IHS) and Principal Component Analysis (PCA), as well as Multiresolution Analysis (MRA) approaches based on wavelets or Laplacian pyramids. While computationally efficient, these methods often introduce spectral distortions due to imperfect separation of spatial and spectral information, which can adversely affect vegetation-related indices and biophysical retrievals [111, 112].

More recently, deep learning-based pansharpening methods have demonstrated superior visual quality and improved quantitative performance. These methods can be broadly classified into three groups: (i) source image concatenation approaches, in which PAN and MS images are directly combined at the input of a neural network; (ii) feature concatenation approaches, where separate subnetworks extract spatial and spectral features prior to fusion to reduce redundancy and distortion; and (iii) feature-domain fusion approaches, which merge complementary PAN and MS information through learned fusion rules [125]. Comprehensive reviews have shown that learning-based methods can outperform classical techniques in terms of spatial detail preservation, but may still suffer from spectral inconsistencies when evaluated for quantitative remote sensing applications [35, 111].

Relevance and Limitations for Biophysical Variable Retrieval

While pansharpening can improve the apparent spatial detail of multispectral imagery, its suitability for biophysical variable retrieval remains debated. LAI estimation relies on physically meaningful spectral relationships, particularly in the red, near-infrared and red-edge bands. Pansharpening modifies these spectral signals through fusion with the PAN band, which may not be spectrally representative of vegetation reflectance and can therefore bias vegetation indices or inversion-based retrievals [112, 118]. As a result, improvements in spatial resolution do not necessarily translate into improved LAI accuracy.

Several studies have highlighted that pansharpened imagery may improve qualitative interpretation and object delineation, but does not consistently improve, and can even degrade, the accuracy of quantitative biophysical products if spectral distortions are not carefully controlled [111]. This concern is particularly relevant in multisensor and learning-based frameworks, where preserving spectral consistency across sensors is critical.

Role of Pansharpening in This Thesis

In the context of this thesis, pansharpening is not treated as a default preprocessing step, but rather as a potential auxiliary technique whose impact must be evaluated empirically. The primary focus remains on preserving physically consistent multispectral information for LAI estimation across sensors. Accordingly, pansharpening is applied only in cases where higher spatial resolution is required for specific analyses, and its impact on spectral fidelity and LAI retrieval performance is explicitly evaluated. This conservative approach is consistent with current best practices in quantitative remote sensing and ensures that any gains in spatial detail do not compromise the reliability of biophysical inference.

2.4 Remote Sensing for LAI Estimation: Evolution of Methods

Leaf Area Index (LAI) quantifies the total one-sided area of leaf tissue per unit ground surface area (m^2/m^2). It serves as a key biophysical parameter for characterizing vegetation canopy structure and function [34]. Accurate spatially continuous LAI maps are essential for agricultural monitoring, carbon cycle modelling and ecosystem assessment. However, field measurement of LAI through destructive harvesting or optical instruments provides only sparse point observations at specific dates [119]. Satellite remote sensing offers the only viable path to wall-to-wall LAI estimation at regional to global scales.

The estimation of LAI from satellite imagery has progressed through four methodological generations, each addressing limitations of its predecessor while introducing new challenges. Figure 2.8 illustrates the general pathway, where spectral reflectance measurements are transformed through increasingly sophisticated methods to produce LAI estimates.

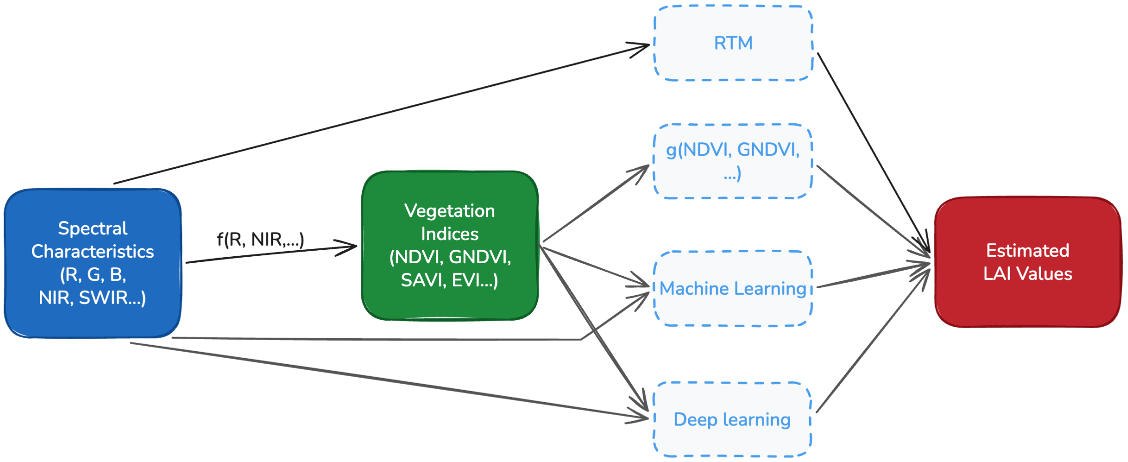


FIGURE 2.8: Schematic overview of methods for deriving Leaf Area Index (LAI) from spectral reflectance measurements.

Physics-Based Radiative Transfer Inversion

The earliest conceptual framework for retrieving biophysical parameters from spectral measurements relies on radiative transfer theory. Radiative transfer models (RTMs) such as PROSAIL simulate canopy reflectance as a forward function of biophysical parameters (LAI, chlorophyll content, leaf angle distribution) and observational geometry [54]. LAI estimation then becomes an inverse problem: finding the parameter set that minimizes the discrepancy between modelled and observed reflectance [109].

However, RTM inversion faces severe practical challenges. The inverse problem is ill-posed: multiple parameter combinations can produce nearly identical spectra. This requires either strong prior constraints or ensemble sampling methods that are computationally prohibitive for operational processing [109]. Computational cost scales poorly with scene size, often requiring hours per satellite image. Furthermore, accuracy depends critically on correct specification of canopy structure parameters (leaf angle distribution, clumping index) that vary spatially and are rarely known *a priori*. These practical limitations motivated the development of simpler empirical approaches.

Vegetation Indices: From Spectral Bands to Biophysical Proxies

Rather than inverting a full physical model, vegetation indices (VIs) exploit the known spectral response of green vegetation through simple algebraic combinations of reflectance bands (the first arrow in Figure 2.8). The Normalized Difference Vegetation Index (NDVI) exploits the strong contrast between high near-infrared reflectance and low red reflectance in healthy vegetation due to chlorophyll absorption [100]. Subsequent indices addressed specific limitations: the Soil-Adjusted Vegetation Index (SAVI) introduced a soil brightness correction factor [49], the Enhanced Vegetation Index (EVI) incorporated blue band reflectance to reduce atmospheric contamination [48] and red-edge indices such as MTCI exploited Sentinel-2's 705 nm and 740 nm bands to extend the dynamic range for high-biomass canopies [40].

Crucially, VIs serve a dual role in LAI estimation. First, they function as biophysical proxies: calibrated empirical relationships (typically exponential or logarithmic transfer functions grounded in Beer-Lambert law attenuation) convert index values directly to LAI estimates without any intermediate modelling step. For example, the relationship $\text{LAI} = -\frac{1}{k}\ln(1 - f(\text{VI}))$ derives from the exponential attenuation of radiation through the canopy, where k is an extinction coefficient and f is an index-specific scaling function [20]. This direct VI-to-LAI pathway offers computational simplicity and physical interpretability.

Second, VIs serve as engineered features for downstream statistical and machine learning models (the VI-to-LAI pathway in Figure 2.8). In this role, the index compresses multi-band spectral information into lower-dimensional representations that are better conditioned for regression. The distinction between these two roles – direct estimation through calibrated functions versus feature engineering for data-driven models – is central to understanding the subsequent evolution of LAI retrieval methods.

Despite their utility, all empirical indices share fundamental limitations: saturation at moderate to high LAI values (NDVI typically plateaus around LAI = 3-5), sensor-specific calibration requirements, region-dependent parametrisation and inability to represent complex non-linear spectral-LAI relationships across diverse agricultural landscapes [48].

Statistical and Machine Learning Regression

Machine learning methods bypass explicit physical modelling by learning empirical mappings from spectral features to LAI through supervised training on labelled data [110]. Early approaches employed shallow models – random forests, support vector regression, Gaussian processes – which demonstrated improved accuracy over single-index calibration curves while maintaining computational efficiency [110]. These models accept either raw spectral bands or vegetation indices (or both) as input features and learn non-linear regression functions through data-driven optimisation.

The key advantage over direct VI-to-LAI transfer functions is that ML regression can combine information from multiple indices and spectral bands simultaneously, capturing interactions that no single index represents. A random forest trained on NDVI, SAVI, EVI and red-edge reflectance jointly exploits complementary spectral information, partially overcoming the saturation limitations of individual indices.

However, these methods still rely on hand-crafted features (vegetation indices, texture metrics) and treat each pixel independently, failing to exploit the spatial structure inherent in multispectral imagery. Furthermore, they require representative training labels, typically derived from field campaigns or the same VI-based calibrations they aim to improve upon.

Deep Learning: End-to-End Spectral-Spatial Mapping

Deep convolutional neural networks represent a fundamental shift in the LAI estimation paradigm. Rather than relying on hand-crafted vegetation indices as intermediate representations, CNNs learn hierarchical feature representations directly from raw spectral bands through end-to-end optimisation [57, 65, 127]. This bypasses the VI computation step entirely, allowing the network to discover spectral-spatial features that may be more informative for LAI prediction than any predefined algebraic band combination.

For LAI estimation specifically, CNNs offer several advantages: (1) learning complex non-linear spectral-spatial relationships without saturation, (2) exploiting spatial context at multiple scales through convolutional receptive fields, (3) end-to-end optimisation from raw reflectance to LAI without intermediate feature engineering and (4) potential for transfer learning across sensors and regions [120].

However, deep learning introduces a critical bottleneck: the requirement for spatially dense training labels. Field-measured LAI is available only at point locations on specific dates, insufficient for training networks that require thousands to millions of labelled pixels. This ground truth problem is addressed in this thesis through an ensemble vegetation index approach that generates pseudo-labels for supervised training while maintaining sensor-agnostic consistency critical for multi-sensor fusion [1] (detailed in Chapter 3, Section 3.1).

2.5 Deep Learning for Semantic Segmentation in LAI Estimation

Dense biophysical parameter estimation from satellite imagery constitutes a semantic segmentation task – assigning a continuous value (LAI) to every pixel in the input image. This section reviews the architectural innovations that enable effective pixel-wise prediction, culminating in the U-Net design employed in this thesis.

2.5.1 U-Net: Encoder-Decoder Architecture for Dense Prediction

The U-Net architecture, introduced by Ronneberger et al. (2015) for biomedical image segmentation [89], has become the de facto standard for dense prediction tasks in both medical imaging [62, 104] and remote sensing [50, 60, 126]. Its defining characteristics address fundamental challenges in pixel-wise regression and classification.

U-Net employs an encoder-decoder structure where the encoder progressively downsamples the input through convolutional and pooling layers, capturing increasingly abstract and spatially coarse features, while the decoder upsamples features back to input resolution through transposed convolutions. This contracting-expanding pathway enables the network to aggregate global context (necessary for understanding scene-level patterns) while recovering fine spatial detail (necessary for pixel-accurate predictions).

The critical innovation of the U-Net architecture is the use of skip connections, which directly concatenate encoder features with decoder features at matching spatial resolutions [89]. These connections serve three purposes: they reintroduce low-level spatial details lost during downsampling, enabling precise localisation; they provide alternative gradient pathways during backpropagation, mitigating vanishing gradients in deep networks; and they fuse coarse semantic information from the decoder with fine-grained spatial detail from the encoder at each resolution level.

These properties are particularly well suited to satellite-based LAI estimation. Agricultural LAI exhibits structure across multiple spatial scales, from sub-metre crop row patterns to field-scale management zones to landscape-level soil and topographic gradients. U-Net encoders capture this hierarchical structure through progressive spatial downsampling, with each pooling layer doubling the receptive field, while skip connections ensure that fine-scale detail is not discarded during reconstruction. Context aggregation is equally relevant, as LAI at a given pixel depends not only on its local spectral signature but on surrounding context. A spectrally bright pixel may indicate high-LAI cropland or bare soil depending on field boundaries and neighbouring vegetation and the contracting path of a U-Net aggregates progressively larger spatial context, with the bottleneck receptive field spanning hundreds of metres. Being fully convolutional, U-Net can process images of arbitrary size and produce spatially aligned output maps, which is advantageous for remote sensing applications where training is performed on small tiles but inference must be applied to full satellite scenes [60].

Atrous Spatial Pyramid Pooling has been proposed as a complementary module for capturing multi-scale context without sacrificing spatial resolution [19], with more recent hybrid approaches further enhancing this capability by integrating self-attention mechanisms [18, 121]. Since agricultural LAI patterns span spatial scales from individual fields to the broader landscape, atrous convolutions offer an efficient means of expanding the receptive field without downsampling or increasing parameter count [123]. For multi-sensor applications, Conditional Batch Normalization provides a mechanism for accommodating sensor-specific feature distributions within a shared encoder-decoder backbone

[24], maintaining unified weights while adapting normalization statistics per input source [83], which has been shown to support effective transfer learning with minimal parameter overhead.

Since its introduction, U-Net and its variants have been widely adopted across remote sensing tasks. In land cover mapping, the architecture has been applied to road extraction [126], building footprint detection [50], crop type classification [91] and multi-resolution land cover segmentation [114]. Beyond static mapping, U-Net-based approaches have been extended to change detection through temporal semantic segmentation of multi-date imagery [23] and to biophysical parameter estimation including canopy height modelling from LiDAR [60] and biomass estimation from SAR [93]. The architecture has also been applied to super-resolution tasks aimed at enhancing the spatial resolution of multispectral imagery [117].

For LAI estimation specifically, several studies have demonstrated U-Net's superiority over shallow regression models and standard CNN classifiers, with reported RMSE improvements of 30-50% relative to vegetation index baselines [82, 120]. However, most prior work has been conducted in single-sensor scenarios, leaving multi-sensor fusion largely unexplored within encoder-decoder frameworks. The present work addresses this gap by extending U-Net to multi-sensor fusion through conditional batch normalization and sensor-specific input projections, as described in Section 2.5.3 and Section 3.2.7.

2.5.2 Dilated Convolutions and Multi-Scale Context

Expanding the receptive field – the spatial extent influencing each output pixel – is critical for capturing large-scale context in remote sensing imagery. Traditional approaches downsample aggressively (reducing spatial resolution) or stack many layers (increasing depth and parameters). **Dilated (atrous) convolutions** [19, 123] provide an elegant alternative – exponentially expanding receptive fields without resolution loss or parameter growth. Recent advances in this domain have further refined these operations, stacking them to achieve hierarchical feature representation that better handles the complexity of dense prediction tasks [116].

A standard discrete convolution with kernel w and input x computes:

$$y[i] = \sum_k w[k] \cdot x[i + k] \tag{2.4}$$

A dilated convolution with dilation rate r inserts $(r - 1)$ zeros (gaps) between kernel weights:

$$y[i] = \sum_k w[k] \cdot x[i + r \cdot k] \tag{2.5}$$

For example, a 3×3 kernel with rate $r = 2$ samples a 5×5 spatial region (with spacing), achieving a receptive field of 5 pixels while retaining only 9 parameters. Rate $r = 4$ yields a 9-pixel receptive field and so on.

For tasks such as LAI mapping from high-resolution imagery, dilated convolutions offer clear advantages by preserving fine-scale spatial information while efficiently capturing broader landscape context. Unlike pooling-based downsampling, dilated convolutions maintain full spatial resolution throughout the network, which is critical for field-level LAI mapping where detail at the scale of tens of metres must be preserved while simultaneously aggregating context at the scale of hundreds of metres. Receptive field growth is achieved efficiently, where a stack of dilated convolutions with exponentially increasing rates ($r = 1, 2, 4, 8, \ldots$) achieves exponential receptive field growth with only linear network depth [123], in contrast to standard convolutions where each additional layer contributes only a small increment to the receptive field. Crucially, dilated convolutions use the same number of parameters as their standard counterparts, making them particularly well suited to remote sensing applications where GPU memory constraints limit batch size and parameter-efficient architectures are therefore essential.

While individual dilated convolutions expand the receptive field at a fixed rate, agricultural scenes contain spatial structures at diverse scales simultaneously – crop rows (meters), field boundaries (tens of meters), landscape heterogeneity (hundreds of meters). Spatial Pyramid Pooling (SPP) [43] addresses this by aggregating features at multiple scales in parallel through pooling operations at different window sizes.

Atrous Spatial Pyramid Pooling (ASPP) [19] adapts SPP using dilated convolutions instead of pooling:

1. Apply multiple dilated convolutions in parallel with different rates (e.g., $r \in \{1, 6, 12, 18\}$);
2. Each branch captures context at a different spatial scale;
3. Concatenate outputs and fuse via 1×1 convolution;
4. Optionally include global average pooling for image-level context.

ASPP enables a network to simultaneously reason across sub-field, field and landscape scales within a single module, improving robustness to varying field sizes and spatial autocorrelation structures [26]. When integrated at the bottleneck of a U-Net, ASPP branches with increasing dilation rates can span receptive fields from the scale of individual crop rows to multi-field landscape extents, providing a principled mechanism for multi-scale context aggregation without sacrificing spatial resolution. Consequently, ASPP modules are frequently utilized in modern hybrid encoder-decoder architectures, serving as a robust bridge between local detail extraction and global context reasoning [121].

2.5.3 Normalization Techniques for Deep Networks

Training deep neural networks requires careful management of internal feature distributions to stabilize gradient flow and accelerate convergence. Batch Normalization [52] revolutionized deep learning by normalizing layer activations, but multi-sensor remote sensing introduces unique challenges that motivate conditional normalization strategies.

Standard Batch Normalization

Batch Normalization (BN) [52] addresses internal covariate shift – the phenomenon where layer input distributions change during training as preceding layers update, forcing each layer to continuously adapt. BN normalizes activations to zero mean and unit variance within each mini-batch:

$$\mathrm{BN}(x) = \gamma \frac{x - \mu_{\mathcal{B}}}{\sqrt{\sigma^2_{\mathcal{B}} + \epsilon}} + \beta \tag{2.6}$$

where:

- $\mu_{\mathcal{B}}, \sigma^2_{\mathcal{B}}$: Mean and variance computed over the current mini-batch $\mathcal{B}$;
- γ, β: Learnable scale and shift parameters (allowing the network to undo normalization if beneficial);
- ϵ: Small constant for numerical stability (10^{-5}).

Batch normalization contributes several properties that are beneficial for training deep networks [52]. By normalising activations at each layer, it prevents gradient explosion and vanishing, enabling the use of higher learning rates and accelerating convergence. It also reduces sensitivity to weight initialisation, as the normalization step compensates for poorly scaled initial activations. The stochasticity inherent in computing statistics over mini-batches introduces implicit regularisation that has been shown to reduce overfitting and the resulting stabilisation of gradient flow has been credited with enabling the successful training of very deep architectures such as ResNets and DenseNets [44, 46].

During inference, batch statistics are unavailable (processing single images). BN uses exponential moving averages of training batch statistics, computed as:

$$\mu_{\text{running}} \leftarrow \alpha \mu_{\text{running}} + (1 - \alpha)\mu_{\mathcal{B}} \tag{2.7}$$

$$\sigma^2_{\text{running}} \leftarrow \alpha \sigma^2_{\text{running}} + (1 - \alpha)\sigma^2_{\mathcal{B}} \tag{2.8}$$

with momentum $\alpha = 0.9$ typically. These running statistics substitute for batch statistics during inference.

Conditional Batch Normalization for Multi-Sensor Learning

In multi-sensor remote sensing, a single network must process inputs from fundamentally different distributions. Sensors differ in spatial resolution, spectral configuration and radiometric response, reflecting sensor-specific calibration procedures and atmospheric correction algorithms. Standard batch normalization computes statistics across all samples in a batch, which forces the network to learn a compromise representation that averages over these distributional differences and has been shown to degrade performance relative to sensor-specific models [28]. This limitation motivates the use of conditional normalization schemes that can adapt internal feature statistics to the source of the input while retaining a shared set of learned weights across the network. Recently, this paradigm has evolved into more flexible dynamic normalization techniques that can adapt not just to fixed domains (sensors), but to varying image content in real-time, further enhancing the model's robustness in heterogeneous environments [14, 47].

Conditional BN [24, 28] generalizes BN by conditioning normalization statistics on auxiliary information – in our case, sensor identity. Instead of global scale/shift parameters (γ, β), we learn sensor-specific parameters $(\gamma^{(s)}, \beta^{(s)})$ for each sensor s:

$$\text{CondBN}(x, s) = \gamma^{(s)} \frac{x - \mu_{\mathcal{B}_s}}{\sqrt{\sigma^2_{\mathcal{B}_s} + \epsilon}} + \beta^{(s)} \tag{2.9}$$

where:

- $s \in \{\text{S2}, \text{L8}\}$: Sensor identifier (passed as metadata with each sample);
- $\mu_{\mathcal{B}_s}, \sigma^2_{\mathcal{B}_s}$: Batch statistics computed *only from sensor s samples*;
- $\gamma^{(s)}, \beta^{(s)}$: Learnable parameters unique to sensor s.

Conditional batch normalization addresses this limitation by maintaining sensor-specific affine parameters while sharing all convolutional weights across the network. Each sensor's feature maps are normalised to their own distribution, avoiding the compromised representations that arise from mixing heterogeneous inputs and the shared convolutional backbone retains the capacity for transfer learning across sensors. The parameter overhead introduced by this scheme is minimal – for a network with N normalization layers and two sensors, conditional batch normalization adds only $2N$ additional scalar parameters, one scale and one shift per sensor per layer, which in practice amounts to well under 1% of total network parameters [28].

Conditional BN was originally developed for neural style transfer [28], where the conditioning variable is artistic style rather than sensor. The core insight – *affine transformation parameters (γ, β) encode domain-specific statistics while convolutional weights encode shared semantic content* – applies equally to multi-sensor fusion [24]. In our context, "style" becomes sensor-specific radiometric and spatial characteristics, while "content" is the underlying LAI spatial pattern.

More recent work has applied conditional normalization to domain adaptation in remote sensing, demonstrating improved cross-region transfer [98] and multi-temporal consistency [29]. However, to our knowledge, this thesis represents the first application of conditional BN specifically to *multi-sensor* LAI estimation where sensors differ fundamentally in resolution and spectral configuration.

Conditional BN requires *single-sensor batches* – all samples in a mini-batch must originate from the same sensor to compute valid batch statistics $\mu_{\mathcal{B}_s}, \sigma^2_{\mathcal{B}_s}$. This constrains batch composition during training (alternating batches between sensors) and complicates distributed training across multiple GPUs where batch sizes are small.

Foundation Models for Earth Observation

A recent paradigm shift in remote sensing has been the emergence of large-scale foundation models pre-trained on massive archives of satellite imagery. Prithvi-EO-2.0 [95], developed jointly by IBM and NASA, exemplifies this trend: a Vision Transformer with masked autoencoder pre-training on 4.2 million Harmonized Landsat-Sentinel-2 (HLS) image sequences at 30 m resolution, available in configurations ranging from 5M to 600M parameters. Prithvi-EO-2.0 incorporates 3D patch embeddings for native spatiotemporal processing and learned temporal-location encodings that condition on acquisition date and geolocation, enabling multi-temporal reasoning across the six spectral bands shared between Landsat and Sentinel-2.

While Prithvi-EO-2.0 has demonstrated strong performance on segmentation tasks (flood mapping IoU 83.1%, wildfire scar IoU 83.2%) and continuous variable regression such as Gross Primary Productivity (R^2 up to 0.88), it has not yet been evaluated for biophysical parameter retrieval tasks such as LAI estimation. This gap is significant: foundation models require fine-tuning on task-specific labelled data and the scarcity of spatially dense, field-validated LAI measurements remains the fundamental bottleneck. Our approach addresses precisely this bottleneck through VI-ensemble pseudo-labels that enable supervised training without field campaigns, while our task-specific U-Net architecture with conditional batch normalization achieves $R^2 = 0.999$ with only 28.7M parameters – a fraction of Prithvi's 600M-parameter flagship variant. The two approaches are complementary rather than competing – foundation model pre-training could provide improved feature representations, while our multi-sensor training strategy and spatial validation methodology would remain essential for fine-tuning any such model for operational LAI retrieval.

Chapter 3

Methodology

This chapter details the complete methodology for multi-sensor LAI estimation, structured as an eleven-stage pipeline (Figure 3.1). The stages span the full workflow from pseudo-ground-truth generation and multi-sensor data integration through preprocessing, augmentation and normalization to model design, training and validation. The modular organisation ensures reproducibility and transparency while allowing individual components to be adapted or extended to other vegetation monitoring tasks. Each stage is described in detail in the sections that follow, with particular attention to the procedures, algorithms and design choices that underpin robust LAI estimation across heterogeneous sensor inputs.

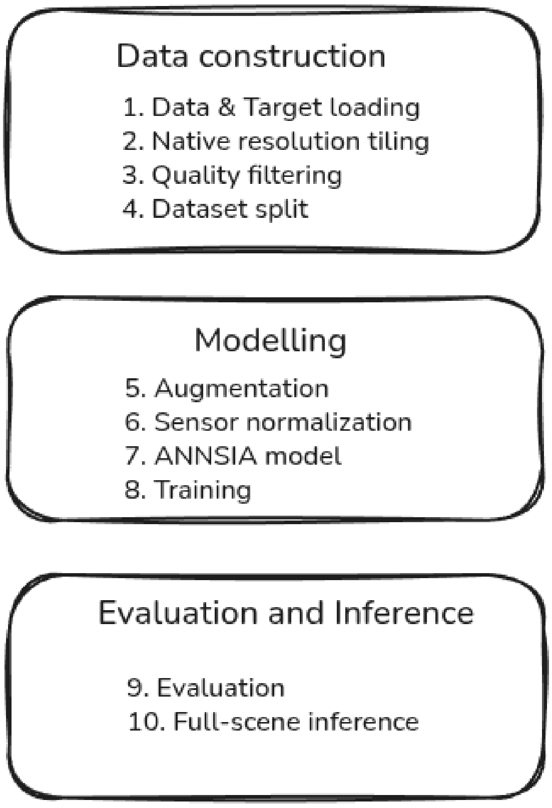


FIGURE 3.1: Overview of the pipeline for multi-sensor LAI estimation, showing the sequence from pseudo-label generation to model training and validation.

3.1 LAI Ground Truth Generation

3.1.1 Motivation and Rationale

The development of supervised deep learning models for LAI estimation faces a fundamental challenge: the scarcity of high-quality ground truth data. Traditional field-based LAI measurements, while accurate, are destructive, labor-intensive and spatially sparse, making them impractical for generating the thousands of training samples required for modern convolutional neural networks. In the Plovdiv study region, no field LAI measurements collected during the 2023 growing season were available.

Existing operational LAI products, such as ESA SNAP's Biophysical Processor and Copernicus Global Land Service (CGLS), offer alternative sources of LAI estimates but present their own limitations. SNAP's neural network implementation of the PROSAIL radiative transfer model tends to extrapolate LAI values beyond the saturation limits of empirical vegetation indices, producing estimates exceeding 6–8 m^2/m^2 in dense agricultural canopies. CGLS products, while validated globally, have coarse spatial resolution (300 m) and limited temporal coverage in our region. Moreover, using operational products as training targets would merely replicate their inherent biases and limitations.

This work adopts an alternative approach: **ensemble vegetation index-based pseudo-labels** as proxy training targets. Rather than treating VI-derived LAI as absolute ground truth, we position them as physically grounded, reproducible proxy targets that enable spatial pattern learning while maintaining transparency about their limitations.

3.1.2 Vegetation Indices Selection

The selection of vegetation indices for the ensemble was guided by three critical requirements:

1. **Sensor agnosticism**: All indices must be computable from spectral bands common to both Sentinel-2 and Landsat 8/9, ensuring consistency across sensors.
2. **Complementary sensitivity**: Indices should exhibit different biophysical responses (e.g., chlorophyll vs. canopy structure, soil adjustment) to capture diverse aspects of vegetation.
3. **Empirical validation**: Prior research demonstrates robust LAI correlation in agricultural systems [67, 71].

Based on these criteria, three vegetation indices were selected for the ensemble – NDVI, GNDVI and SAVI. These indices were chosen because they provide complementary sensitivity to different canopy properties while remaining computable from spectral bands available on both Sentinel-2 and Landsat 8/9 sensors.

NDVI captures the overall vegetation vigor and canopy density signal, GNDVI enhances sensitivity to chlorophyll concentration and internal canopy structure, while SAVI mitigates soil background effects in sparsely vegetated conditions. The combination of these indices enables a more robust representation of vegetation status across a wide LAI range.

Table 3.1 summarizes the spectral formulas of chosen and sensor-specific band mappings for Sentinel-2 and Landsat 8.

TABLE 3.1: Vegetation index formulas and sensor-specific band mappings.

Index	Formula	Arguments	S2 Bands	L8/9 Bands	Saturation LAI
NDVI	$\frac{\rho_{\text{NIR}}-\rho_{\text{Red}}}{\rho_{\text{NIR}}+\rho_{\text{Red}}}$	NIR, Red	B08, B04	B5, B4	∼3–4
GNDVI	$\frac{\rho_{\text{NIR}}-\rho_{\text{Green}}}{\rho_{\text{NIR}}+\rho_{\text{Green}}}$	NIR, Green	B08, B03	B5, B3	∼4–5
SAVI	$\frac{(\rho_{\text{NIR}}-\rho_{\text{Red}})(1+L)}{\rho_{\text{NIR}}+\rho_{\text{Red}}+L}$	NIR, Red	B08, B04	B5, B4	∼3–4

3.1.3 VI-to-LAI Transfer Functions

For all three selected indices we use the **exponential (Beer-Lambert) transfer function**, which models LAI as the optical thickness required to attenuate incident radiation according to:

$$\text{LAI} = -\frac{1}{k}\ln(1-\text{VI}) \tag{3.1}$$

where:

- VI is the vegetation index value, clipped to $[0, 0.95]$ prior to transformation (raw normalized-difference indices range from -1 to $+1$, but negative values indicate non-vegetated surfaces and are clamped to zero);
- k is the empirical extinction coefficient, calibrated per index;
- The logarithmic relationship enforces an asymptotic saturation behaviour.

The extinction coefficients were adopted from established literature calibrations for agricultural crops:

- **NDVI**: $k = 0.60$ [20];
- **GNDVI**: $k = 0.60$ [38];
- **SAVI**: $k = 0.60$ [49].

From this point onward, LAI values derived via this exponential transformation will be referred to as **NDVI-exp**, **GNDVI-exp**, and **SAVI-exp**, respectively. This convention is used consistently in subsequent analyses, figures and comparisons.

The Beer-Lambert formulation has strong theoretical grounding in radiative transfer theory: the attenuation of radiation through a vegetation canopy follows exponential decay proportional to cumulative leaf area. While simplified (assumes random leaf distribution, ignores multiple scattering), it provides a physically interpretable link between spectral reflectance and canopy structure.

3.1.4 Ensemble Aggregation Strategy

Individual vegetation indices exhibit complementary strengths and weaknesses. NDVI saturates earlier but responds strongly to chlorophyll content; GNDVI extends the dynamic range; SAVI mitigates soil background influence. To leverage this complementarity and reduce method-specific biases, **ensemble averaging** is employed:

$$\mathrm{LAI}_{\mathrm{ensemble}} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{LAI}_{\mathrm{method}_i} \tag{3.2}$$

where $N = 3$ (NDVI, GNDVI, SAVI exponential methods) and $\mathrm{LAI}_{\mathrm{method}_i}$ is the LAI estimate from the i-th vegetation index.

The inter-method variability provides a pixel-wise estimate of uncertainty via the standard deviation across methods:

$$\sigma_{\mathrm{LAI}} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (\mathrm{LAI}_{\mathrm{method}_i} - \mathrm{LAI}_{\mathrm{ensemble}})^2} \tag{3.3}$$

High inter-method standard deviation ($\sigma > 0.5\ \mathrm{m}^2/\mathrm{m}^2$) indicates:

- Saturation onset (methods diverging due to different saturation points)
- Soil/atmosphere contamination (affecting indices differently)
- Mixed pixels (vegetation/non-vegetation boundaries)

The resulting uncertainty metric is stored alongside the ensemble mean and serves three purposes within the broader pipeline. First, it enables training sample quality filtering by excluding tiles whose uncertainty exceeds 1.0 $\mathrm{m}^2/\mathrm{m}^2$. Second, it supports result interpretation by identifying spatial regions where LAI estimates are considered less reliable. Third, it provides a foundation for future weighting schemes in which uncertain samples may be down-weighted during model training.

Preliminary experiments demonstrated that ensemble averaging offers several advantages over any single spectral index. Random errors in individual methods partially cancel through error averaging, while the use of indices with differing saturation thresholds extends the effective dynamic range and mitigates overall saturation. The ensemble is furthermore less sensitive to single-method failures such as those arising from soil contamination. A CNN trained on ensemble targets can subsequently outperform individual vegetation indices, with a demonstrated significant RMSE reduction (see Chapter 4.2), by learning spatial patterns that extend beyond pixel-level spectral relationships.

3.1.5 Justification as Proxy LAI Target

The use of VI-ensemble LAI estimates as training targets requires careful justification, particularly given the **circular nature** of training a CNN on VI-derived labels to ultimately produce LAI maps. This work does NOT claim to produce absolute LAI retrievals validated against real-world field sampling. Instead, we position the methodology as **operational LAI estimation using VI-ensemble pseudo-labels**, with the following justifications:

1. Consensus Across Established Methods

The ensemble aggregates three independent VI-to-LAI formulations, each with decades of empirical validation in agricultural systems [38, 49, 100]. Convergence among methods increases confidence in the resulting LAI estimate.

2. Physical Basis

The Beer-Lambert transfer function is grounded in radiative transfer theory, not arbitrary curve fitting. While simplified, it captures the fundamental exponential attenuation of light through vegetation canopies.

3. Sentinel-2 Validation Heritage

ESA's validation of Sentinel-2 LAI products reports RMSE < 1.0 m^2/m^2 for agricultural systems [9], confirming that VI-based approaches can achieve acceptable accuracy in cropland contexts similar to our study region.

4. Independence from SNAP

Unlike using SNAP LAI as training targets (which would replicate SNAP's neural network), VI-ensemble labels are independently derived. Comparison with SNAP provides an orthogonal validation perspective (Section 4.4.2).

5. Spatial Consistency

VI-based LAI estimates exhibit strong spatial coherence, capturing field boundaries, crop types and phenological gradients. Even if absolute values differ from field measurements, the spatial patterns provide valuable training signal for the CNN to learn texture, context and vegetation structure.

6. Addressing Multiple Crops and Phenology

Our study region contains diverse crop types (wheat, maize, sunflower, vineyards) at varying phenological stages in September. VI-ensemble averaging mitigates crop-specific biases better than single-method approaches.

3.1.6 Limitations

We acknowledge the following fundamental limitations of the VI-ensemble proxy target approach:

1. Circular Ground Truth

The most critical limitation is that the CNN is trained on VI-derived LAI labels and validated by comparing predictions to those same VI-ensemble estimates or their derivatives. This circularity means the model's task is to spatially interpolate and smooth VI-ensemble LAI rather than retrieve absolute LAI independent of vegetation indices. Chapter 4 addresses this concern through a series of complementary assessments. The CNN output is compared against SNAP as an independent physics-based reference and spatial coherence analysis is conducted to determine whether the CNN improves the smoothness of LAI estimates. Cross-region and temporal generalization experiments further evaluate whether the model learns transferable spatial patterns, while per-sensor performance evaluation assesses whether multi-sensor fusion provides measurable added value.

2. Saturation Inheritance

VI-based methods saturate at moderate-to-high LAI values in the range of 3-5 m^2/m^2 and a CNN trained on such targets will inevitably inherit this saturation limitation, leading to potential underestimation in dense forests or mature crop canopies. This work mitigates the concern by focusing on agricultural landscapes during the growing season, where LAI typically remains below 5 m^2/m^2 and saturation is consequently less problematic.

3. Single-Season Calibration

Training data are derived from September 2023 scenes and the extinction coefficients and ensemble averaging may not generalize beyond this context. Seasonal variation such as senescence or winter dormancy, differences in vegetation type across forests or grasslands and shifts in climatic region all represent potential sources of failure. Temporal and cross-region generalization are nonetheless tested using May 2024 data acquired over Burgas, a landscape distinct from the training region, with results presented in Section 4.3.1 confirming cross-region transferability and temporal stability across an eight-month gap.

3.2 Data Pipeline

3.2.1 Data Loading and Sensor Configuration

The methodology employs surface reflectance products from both sensors. Sentinel-2 Level-2 products are obtained from the Copernicus Data Space, while Landsat 8/9 Collection 2 Level-2 products are sourced from the USGS Earth Explorer. Both product levels apply atmospheric correction to convert top-of-atmosphere reflectance to bottom-of-atmosphere surface reflectance, removing the effects of atmospheric scattering and absorption. This preprocessing step is critical for ensuring consistent spectral comparisons across multiple acquisition dates and sensor platforms.

For Sentinel-2, all bands with a native resolution of 20 m or better are retained, comprising ten bands in total. The 60 m coastal, aerosol and cirrus bands (B01, B09 and B10) are excluded, while the four red-edge bands (B05, B06, B07 and B8A) unique to Sentinel-2 are preserved for their enhanced sensitivity to chlorophyll content and vegetation structure. For Landsat 8/9, all six 30 m VNIR-SWIR bands (B2-B7) are used, excluding the 15 m panchromatic band and the thermal bands. The spectral coverage of these bands aligns with the visible, NIR and SWIR regions present in Sentinel-2. Across both sensors, six common spectral regions comprising Blue, Green, Red, NIR, SWIR1 and SWIR2 form a sensor-agnostic subset that ensures equivalent information contribution from each platform, while the additional red-edge bands of Sentinel-2 provide complementary spectral detail where available. Table 3.2 summarizes the selected bands and their spectral correspondence.

TABLE 3.2: Spectral band selection for Sentinel-2 and Landsat 8/9.

Spectral Region	**S2 Band**	**S2 λ (nm)**	**L8/9 Band**	**L8/9 λ (nm)**
Blue	B02	490	B2	482
Green	B03	560	B3	561
Red	B04	665	B4	655
Red-Edge 1	B05	705	–	–
Red-Edge 2	B06	740	–	–
Red-Edge 3	B07	783	–	–
NIR	B08	842	B5	865
NIR Narrow	B8A	865	–	–
SWIR 1	B11	1610	B6	1609
SWIR 2	B12	2190	B7	2201
Total Bands	**10 bands**		**6 bands**	

Raw digital numbers in the satellite products are converted to physical reflectance values prior to processing. For Sentinel-2, integer reflectance values stored in the range 0-10,000 are divided by 10,000 to yield floating-point reflectance in the range 0-1. For Landsat 8/9, scaled reflectance values are obtained by applying the gain and offset parameters provided in the USGS metadata. All bands are stored as `float32` during processing and subsequently converted to `float16` for efficient storage in the final dataset (Section 3.2.4).

3.2.2 Tiling Strategy at Native Resolution

Unlike conventional multi-sensor fusion approaches that resample all inputs to a common spatial grid, typically at 10 m, the present methodology preserves native sensor resolution throughout the training pipeline. Each sensor is tiled at its original pixel spacing, with tile dimensions adjusted to cover approximately the same geographic area. This design choice offers several advantages. Preserving native resolution avoids the resampling artifacts such as smoothing and aliasing that can degrade fine-scale spatial information and it respects each sensor's point spread function, which reflects the intrinsic spatial response of the underlying optical system and is distorted when resampling is applied. Furthermore, the CNN architecture is designed to accommodate different input resolutions through sensor-specific pathways (Section 3.2.7), treating resolution diversity as a source of complementary information rather than a technical limitation to be resolved.

Table 3.3 summarizes the tile dimensions for each sensor, selected to achieve approximately 1.58 km × 1.58 km ground coverage.

TABLE 3.3: Tile size configuration for native resolution processing.

Sensor	**Resolution**	**Tile Size**	**Coverage (km)**	**Coverage (km^2)**	**Pixels**
Sentinel-2	10 m	158×158	1.58 × 1.58	2.50	24,964
Landsat 8/9	30 m	53×53	1.59 × 1.59	2.53	2,809

The ground coverage mismatch is negligible (2.50 km^2 vs. 2.53 km^2, <1.2% difference), ensuring both sensors capture equivalent spatial context for field-scale agricultural monitoring.

Full scenes are subdivided into non-overlapping tiles using a sliding window with a stride equal to the tile size. Adjacent tiles share no pixels, which prevents spatial autocorrelation leakage between the training, validation and test sets, a property that is critical for the spatial block validation strategy employed Section 3.2.4. Partial tiles arising at scene boundaries are discarded if they fail to meet a minimum size threshold of 90% of the nominal tile dimensions. Throughout this process, each tile retains its coordinate reference system and geotransform metadata, enabling spatial validation and full-scene reconstruction during inference.

3.2.3 Quality Filtering and Masking

Cloud contamination is a pervasive challenge in optical remote sensing. Both sensors provide pixel-level quality assessment layers to address this issue.

In addition, some scenes contain black pixels corresponding to NoData values. These arise where the satellite acquisition footprint does not completely overlap the rectangular raster grid used for data storage. Pixels outside the valid observation area are assigned null values and appear as black regions in the imagery.

For Sentinel-2, the Sen2Cor processor generates a 20 m Scene Classification Layer (SCL) raster with twelve classification categories [70]. Pixels classified as "Cloud High Probability", "Cloud Medium Probability", "Cloud Shadows", "Thin Cirrus", "Snow or Ice", "Saturated or Defective", "Dark Area Pixels", "Unclassified" and "No-data" are masked. Valid vegetation pixels correspond to SCL classes "Vegetation", "Not Vegetated" and "Water", after manual inspection confirmed their agricultural relevance.

For Landsat 8/9, the Collection 2 QA_PIXEL band uses bit-packed flags to indicate pixel reliability instead of simple classification labels. Conservative masking is applied to pixels with high cloud confidence, high cloud shadow confidence, high cirrus confidence, saturated values in any band and fill values representing no data.

After pixel-level masking, tiles are evaluated for overall quality:

$$\text{Valid Fraction} = \frac{\text{Count}(\text{valid pixels})}{\text{Total tile pixels}} \tag{3.4}$$

Tiles are retained if:

- Valid Fraction ≥ 0.95 (95% cloud-free coverage);
- LAI ensemble mean $\in [0, 10]$ m^2/m^2 (excludes spurious/invalid LAI estimates);
- LAI ensemble standard deviation < 1.5 m^2/m^2 (excludes high-uncertainty regions).

These thresholds balance data retention with quality assurance, prioritizing high-confidence samples for training.

Masked pixels within retained tiles are assigned NaN values and are excluded from loss computation via explicit masking in the loss function. This approach prevents contaminated or unreliable spectral values from biasing gradient updates and allows training to proceed on partially cloudy tiles without discarding entire samples.

3.2.4 Dataset Generation and Spatial Splitting

Processed tiles are serialized into a structured dataset hierarchy, with each `.npy` file containing three components. The reflectance array is a multi-channel tensor with spatial dimensions matching the native sensor resolution, comprising ten channels for Sentinel-2 and six for Landsat 8/9. The ensemble mean LAI and its associated uncertainty are stored as separate single-channel arrays, both expressed in units of m^2/m^2.

```
dataset_root/
|-- train/
|   |-- reflectance/  # Sensor bands (float16, shape: [C, H, W])
|   |   |-- S2_tile_0001.npy
|   |   |-- L8_tile_0001.npy
|   |   `-- ...
|   `-- target/       # LAI ensemble mean/std (float16)
|       |-- lai_mean_0001.npy
|       |-- lai_std_0001.npy
|       `-- ...
|-- val/   (same structure)
`-- test/  (same structure)
```

Storage in `float16` half-precision reduces the disk footprint by 50% relative to `float32` while maintaining sufficient numerical precision for reflectance values to four significant digits and LAI to a resolution of 0.001 m^2/m^2. The resulting per-tile storage requirements are approximately 498 KB for a Sentinel-2 reflectance tile of dimensions $10 \times 158 \times 158$ and 34 KB for a Landsat 8/9 tile of dimensions $6 \times 53 \times 53$. The corresponding LAI target arrays, storing both the ensemble mean and uncertainty, occupy approximately 100 KB per Sentinel-2 tile and 11 KB per Landsat 8/9 tile.

Geographic disjointness is implemented through a 4×4 spatial grid partitioning of the full scene into sixteen equal blocks arranged in four rows and four columns (Figure 3.2). Blocks are randomly assigned to training, validation and test splits using a fixed seed, with a target 70/15/15 ratio. This configuration ensures that all tiles within the same geographic block are assigned to the same split, and that each split samples a distinct subregion of the scene. Test performance therefore reflects true out-of-sample prediction rather than interpolation within a spatially autocorrelated neighbourhood.

This spatial block splitting strategy is motivated by the noted problem of spatial autocorrelation in geospatial machine learning, where random or pixel-level splits can lead to too much optimistic performance estimates due to information leakage between neighbouring samples. Prior works have shown that models evaluated under non-spatially aware splitting schemes often fail to generalize when deployed to geographically distinct regions, as they implicitly exploit local spatial dependencies rather than learning robust, transferable representations [59]. Spatially disjoint validation protocols, such as block-wise cross-validation, have therefore been recommended as a more realistic evaluation framework for remote sensing and environmental modelling tasks. In particular, Roberts et al. in 2017

[87] demonstrate that ignoring spatial structure in cross-validation can significantly inflate predictive performance, especially in highly autocorrelated environmental datasets. Similarly, more recent analyses in Earth observation literature (e.g. [102]) emphasize that spatial blocking better approximates operational deployment conditions, where models must generalize to unseen geographic regions rather than interpolate within the same scene. By enforcing strict spatial separation between training and evaluation data, the adopted approach reduces spatial leakage and provides a more conservative and realistic estimate of generalization performance.

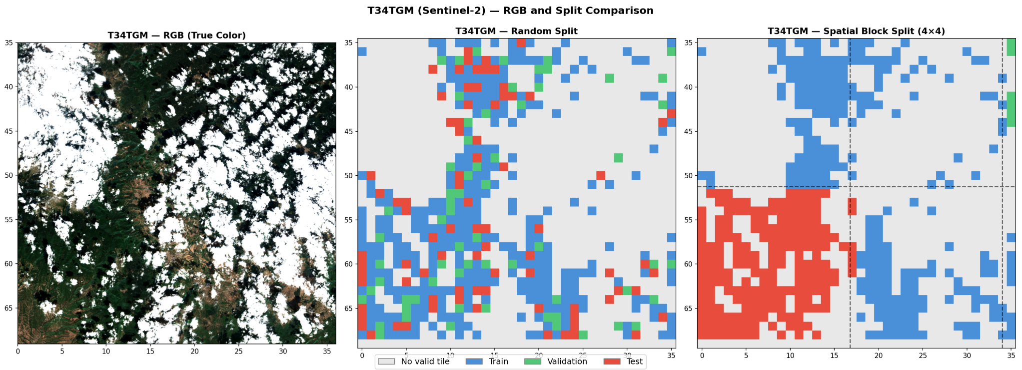


FIGURE 3.2: Cropped detail from a Sentinel-2 image illustrating random and spatial block split strategies.

3.2.5 Data Augmentation

Training data are augmented using a set of eight geometric transformations to improve model robustness and data efficiency, as summarised in Table 3.4.

TABLE 3.4: Geometric augmentation scheme applied during training.

Aug. ID	Transformation	Equivalent View
0	Original (no transform)	0°
1	Rotate 90° clockwise	90°
2	Rotate 180°	180°
3	Rotate 270° clockwise	270°
4	Horizontal flip	Mirror X-axis
5	Vertical flip	Mirror Y-axis
6	Horizontal flip + Rotate 90°	Diagonal flip
7	Vertical flip + Rotate 90°	Anti-diagonal flip

All transformations are applied on-the-fly during batch loading rather than being pre-computed, which avoids any increase in storage requirements.

Reflectance and LAI target images are transformed identically through synchronised augmentation, and because only 90° increments and axis-aligned flips are used, spectral values are preserved exactly without interpolation. A transformation is selected at random per training iteration to ensure variability across epochs.

The choice of this augmentation scheme is motivated by the geometric properties of the target domain. Agricultural fields have arbitrary orientations with respect to the sensor flight path and the CNN should therefore predict LAI independently of cardinal direction. Similarly, vegetation patterns exhibit no preferred left-right or top-bottom asymmetry, making flip invariance a natural constraint to enforce. Beyond these domain-specific considerations, the eightfold augmentation effectively multiplies the number of apparent training samples without requiring additional data acquisition and acts as a regulariser by compelling the model to generalise across spatial orientations, thereby reducing the risk of overfitting.

Spectral augmentation, such as brightness or contrast adjustments and channel-wise perturbations, was deliberately omitted. Vegetation indices and LAI estimates depend on precise spectral ratios and perturbing individual bands without recalculating the corresponding LAI targets would introduce inconsistencies between inputs and labels. Geometric augmentation was therefore deemed sufficient for regularisation while preserving the physical integrity of the spectral-LAI relationships.

For multi-sensor experiments, an additional augmentation strategy addresses the dataset imbalance arising from the difference in ground swath between the two sensors. Landsat 8/9 tiles outnumber Sentinel-2 tiles by approximately four to one and without correction, training batches expose the model disproportionately to Landsat data, leading to biased feature learning. To mitigate this issue, a sensor-aware sampling strategy is applied, increasing the sampling rate of Sentinel-2 tiles during training to balance sensor exposure at the batch level. This approach is consistent with domain-balanced training strategies commonly used in multi-source learning settings [81].

3.2.6 Normalization

Raw surface reflectance values vary across sensors due to differences in spectral response functions, radiometric calibration coefficients, atmospheric correction algorithms and the typical vegetation reflectance distributions associated with each band [101]. These discrepancies must be addressed before multi-sensor training, as deep neural networks train most efficiently when inputs are centred around zero with consistent magnitude. Raw reflectance values spanning the 0-1 range exhibit markedly different distributions across bands, with NIR reflectance typically in the range 0.3-0.6 and SWIR reflectance in the range 0.1-0.3, leading to uneven gradient magnitudes during optimisation. Z-score normalization maps each sensor's reflectance distribution to a common standardised scale, facilitating shared representation learning in the encoder despite inter-sensor spectral differences, while its linear nature preserves the relative ordering and spectral contrasts that are critical for vegetation index computation, in contrast to non-linear alternatives such as histogram equalisation.

To enable effective multi-sensor fusion, sensor-specific Z-score normalization is applied:

$$x_{\text{norm}}^{(s,b)} = \frac{x_{\text{raw}}^{(s,b)} - \mu^{(s,b)}}{\sigma^{(s,b)}} \tag{3.5}$$

where:

- $x_{\text{raw}}^{(s,b)}$: Raw reflectance for sensor s and band b;
- $\mu^{(s,b)}$: Mean reflectance computed over the training set;
- $\sigma^{(s,b)}$: Standard deviation over the training set;
- $x_{\text{norm}}^{(s,b)}$: Normalized reflectance (zero mean, unit variance).

Normalization statistics are computed *separately* for each sensor and each spectral band from the training split only, then applied consistently to validation and test sets.

3.2.7 Model Architecture

The core of this methodology is the Adaptive Neural Network for Satellite Image Analysis (ANNSIA), a multi-sensor U-Net with Conditional Batch Normalization and Atrous Spatial Pyramid Pooling designed to process inputs from multiple sensors at their native resolutions while learning shared representations for LAI prediction.

Figure 3.3 illustrates the complete network, including sensor-specific input projections, the shared encoder–decoder backbone, the ASPP-enhanced bottleneck, and conditional batch normalization layers. This overview serves as a roadmap: the following subsections discuss each component in detail.

Multi-Sensor U-Net Backbone

The U-Net architecture [89] is a fully convolutional encoder-decoder network whose design features make it particularly suitable for dense LAI prediction. The model progressively downsamples the input through an encoder to capture hierarchical contextual features, then upsamples through a symmetric decoder to reconstruct spatial resolution. Skip connections concatenates encoder features with decoder features at matching resolutions, preserving fine-scale spatial detail. Because the network is fully convolutional and contains no fully connected layers, it can operate on arbitrary input sizes (though fixed patch sizes are used during training for batching consistency).

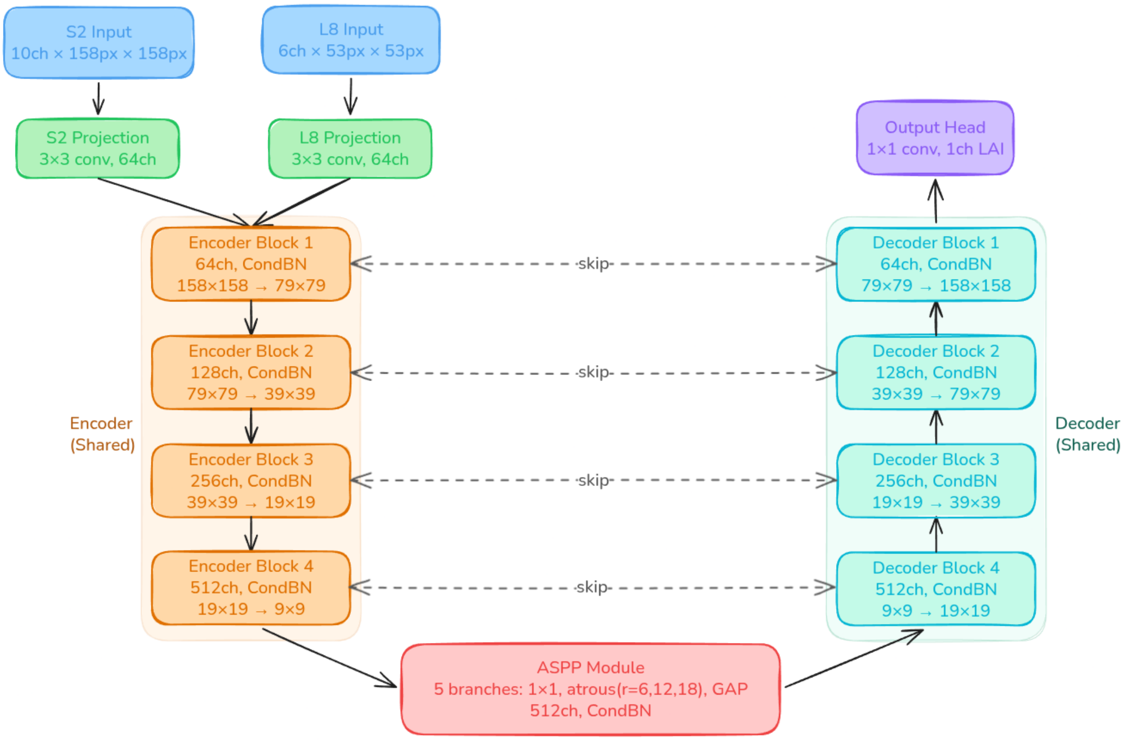


FIGURE 3.3: The architecture of ANNSIA.

U-Net architecture with sensor-specific input projections, shared encoder-decoder, ASPP module at the bottleneck and conditional batch normalization throughout. Skip connections (dashed arrows) concatenate encoder features with decoder features at matching spatial resolutions. Both Sentinel-2 and Landsat 8/9 inputs enter through separate projection layers but share all subsequent encoder-decoder weights.

Sensor-Specific Input Projections: To accommodate different input channel counts and spatial dimensions, the architecture employs sensor-specific input projection layers:

- **Sentinel-2**: 3×3 convolution mapping 10 input channels $\rightarrow$ 64 feature channels;
- **Landsat 8/9**: 3×3 convolution mapping 6 input channels $\rightarrow$ 64 feature channels.

After this projection step, data from both sensors are processed by a shared encoder-decoder backbone with identical weights, enabling transfer learning and sensor-invariant feature extraction.

Encoder Architecture: The encoder consists of four downsampling blocks that progressively increase the feature dimensionality while reducing spatial resolution. Each block follows the same structure (Figure 3.4):

1. Two 3×3 convolutions (ReLU activation);
2. Conditional Batch Normalization (sensor-specific parameters, Section 3.2.7);
3. 2×2 max pooling (stride 2) for spatial downsampling.

The number of feature channels increases across the encoder as:

$$64 \rightarrow 128 \rightarrow 256 \rightarrow 512.$$

For an input Sentinel-2 patch of size 158×158, the spatial resolution progression is as follows:

$$158 \times 158 \rightarrow 79 \times 79 \rightarrow 39 \times 39 \rightarrow 19 \times 19 \rightarrow 9 \times 9.$$

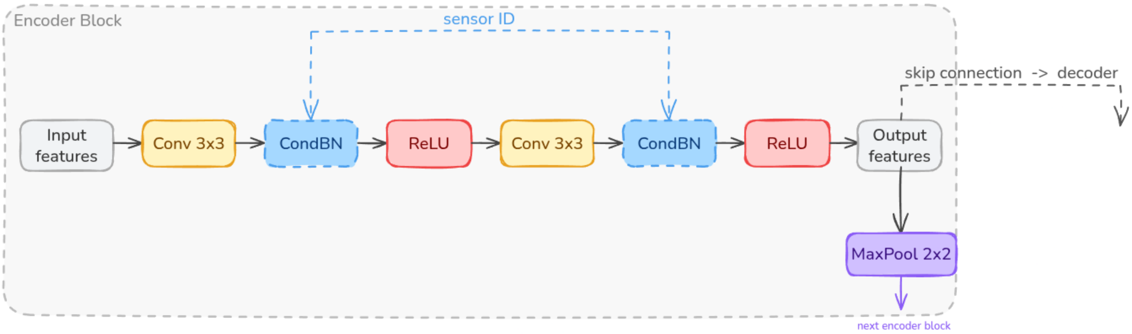


FIGURE 3.4: Architecture of the encoder block.

Bottleneck: At the deepest level of the U-Net, the feature maps have the lowest spatial resolution and the highest semantic abstraction, forming the *bottleneck* between encoder and decoder. Here, an **Atrous Spatial Pyramid Pooling (ASPP)** module is applied to capture multi-scale contextual information before upsampling begins. ASPP enables the network to aggregate features relevant across fine, medium and coarse spatial scales, which is important for accurate LAI prediction. The detailed architecture and operation of ASPP are described below.

Decoder Architecture: The decoder mirrors the encoder with four upsampling blocks, each progressively recovering spatial resolution while integrating high-resolution features from the encoder through skip connections. Each block consists of:

1. 2×2 transposed convolution (stride 2) for spatial upsampling;
2. Concatenation with encoder skip connection features;
3. Two 3×3 convolutions (ReLU activation);
4. Conditional Batch Normalization.

Feature channel progression after concatenation is:

$$1024 \rightarrow 512 \rightarrow 256 \rightarrow 128 \rightarrow 64.$$

Output Head: The final 1×1 convolution maps the 64 feature channels to a single output channel representing LAI, followed by ReLU activation to enforce non-negativity ($\text{LAI} \geq 0$).

Parameter Count: The network contains approximately **28.7 million trainable parameters**, distributed as follows:

- Sensor-specific input projections: 2 × 5.8K = 11.6K;
- Shared encoder-decoder: 28.6M;
- Conditional BN parameters: 102K (sensor-specific scales and biases).

Atrous Spatial Pyramid Pooling (ASPP)

Agricultural LAI exhibits multi-scale spatial patterns. Fine-scale variability arises within fields (e.g., crop rows and irrigation gradients), medium-scale variability corresponds to field boundaries & crop-type transitions and coarse-scale variability reflects regional land patterns and topographic effects. To capture these multi-scale features, we employ the Atrous Spatial Pyramid Pooling (ASPP) module [19], which aggregates features at multiple receptive field scales using atrous (dilated) convolutions, capturing context without reducing spatial resolution.

ASPP is inserted at the encoder bottleneck, where spatial resolution is lowest and semantic abstraction is highest. It consists of five parallel branches (Figure 3.5):

1. 1×1 convolution for point-wise feature extraction;
2. 3×3 atrous convolution, rate $r = 6$;
3. 3×3 atrous convolution, rate $r = 12$;
4. 3×3 atrous convolution, rate $r = 18$;

5. global average pooling followed by 1×1 convolution and bilinear upsampling to capture image-level context.

All branches output 256 channels. Outputs are concatenated ($5 \times 256 = 1280$ channels), then fused via 1×1 convolution back to 512 channels.

An atrous 3×3 convolution with rate r inserts $r - 1$ zeros between kernel weights, expanding the receptive field without increasing parameters:

- Standard conv: receptive field = 3 pixels;
- Atrous conv, $r = 6$: receptive field = $3 + 2(6 - 1) = 13$ pixels;
- Atrous conv, $r = 12$: receptive field = 25 pixels;
- Atrous conv, $r = 18$: receptive field = 37 pixels.

At the bottleneck resolution (e.g., 9×9 for Sentinel-2), these receptive fields translate to:

- $r = 6$: $13 \times 16 = 208$ m context (sub-field scale);
- $r = 12$: $25 \times 16 = 400$ m context (field scale);
- $r = 18$: $37 \times 16 = 592$ m context (multi-field scale).

Here, each bottleneck pixel represents approximately $158/9 \approx 17.6$ m ≈ 16 m at the original 10 m input resolution, after four pooling layers with stride 2.

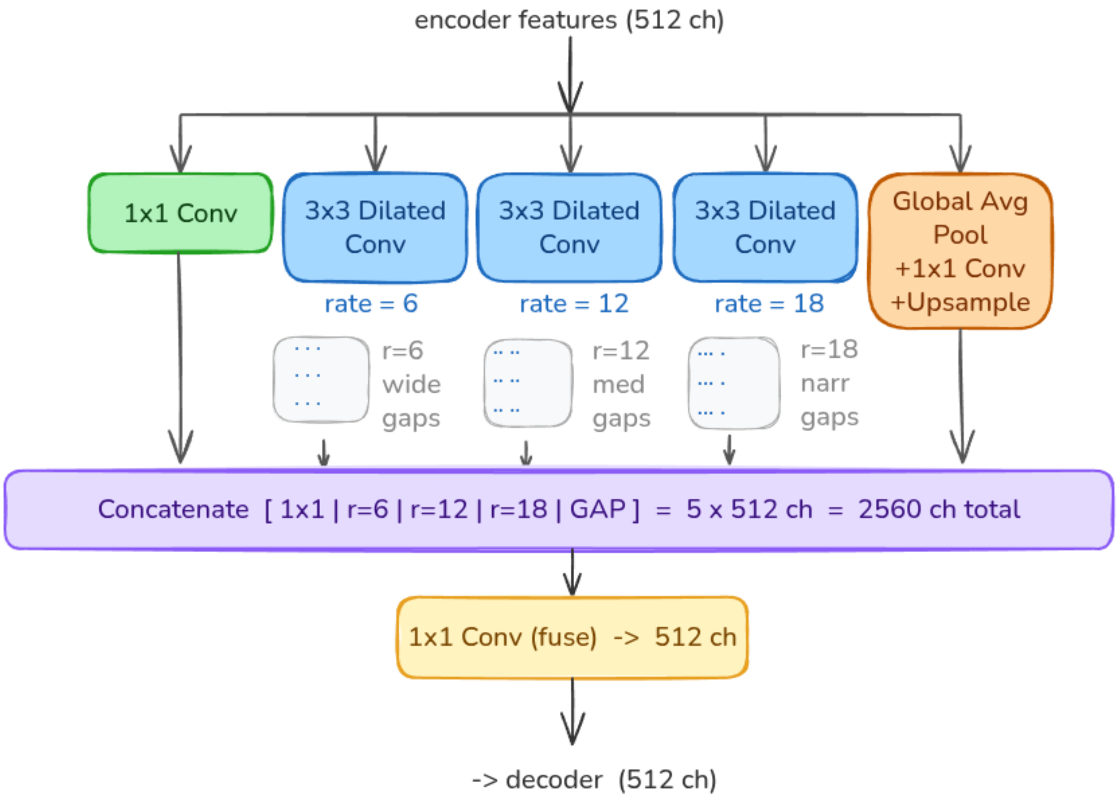


FIGURE 3.5: Architecture of the ASPP.

Crucially, ASPP incorporates **Conditional Batch Normalization** in each branch, enabling sensor-specific feature statistics while sharing convolutional weights.

Conditional Batch Normalization

Although all input bands are Z-score normalized (Section 3.2.6), Sentinel-2 and Landsat 8/9 still exhibit different internal feature distributions within the CNN. These differences arise from several factors, including the different input resolutions (158×158 vs. 53×53 input patches), different spectral information content (10 bands vs. 6 bands) and different spatial frequency characteristics of the sensors.

Standard batch normalization computes global statistics across all training samples, forcing the network toward a compromise representation. To address this, conditional batch normalization is adopted here, allowing sensor-specific normalization statistics while sharing all convolutional weights across the network, as motivated in the background discussion of Section 2.5.3. Each normalization layer maintains two sets of affine parameters, one per sensor, enabling the model to preserve specialised feature distributions for each sensor while retaining a unified convolutional representation throughout the encoder-decoder backbone (Figure 3.6).

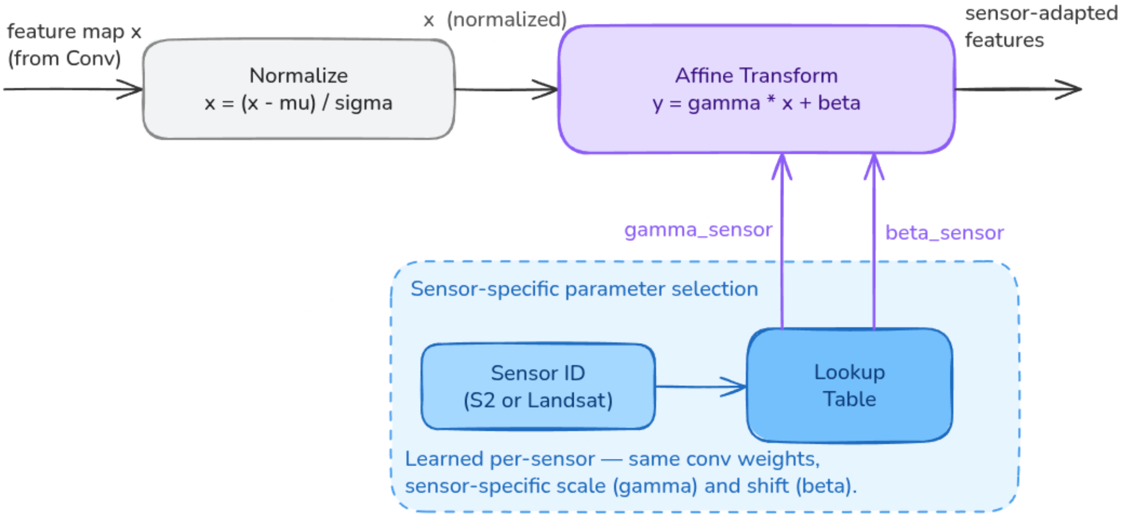


FIGURE 3.6: The mechanism of Conditional Batch Normalization.

Training and Inference

During training, each batch contains samples from a single sensor (alternating batches between Sentinel-2 and Landsat 8/9). The sensor identifier is passed as metadata to all CondBN layers, which select the appropriate sensor-specific normalization parameters $(\gamma^{(s)}, \beta^{(s)})$ and update the sensor-specific batch statistics.

During inference, the sensor identifier is provided with each input tile. The model uses exponential moving averages of the batch statistics computed separately for each sensor during training.

Conditional BN therefore enables the network to normalize Sentinel-2 and Landsat 8/9 features independently, accommodating their different statistical distributions, while simultaneously learning sensor-invariant convolutional filters through shared weights and allowing sensor-specific feature scaling. This design facilitates effective multi-sensor fusion without forcing a single normalization scheme across sensors.

3.2.8 Training Strategy

Optimizer and Learning Rate

Training is performed using the AdamW optimiser [69], which extends Adam with decoupled weight decay to improve generalisation. A base learning rate of $\eta = 1 \times 10^{-3}$ is used alongside a weight decay of $\lambda = 1 \times 10^{-5}$ and standard Adam momentum parameters of $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate follows a cosine annealing schedule with periodic warm restarts [68], allowing the optimiser to escape local minima through cyclical learning rate reductions. AdamW's per-parameter adaptive learning rates stabilise training in deep networks while the decoupled weight decay term acts as an L2 regulariser, penalising large weights without interfering with the adaptive gradient scaling.

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min}) \left(1 + \cos\left(\frac{T_{\text{cur}}}{T_i}\pi\right)\right) \tag{3.6}$$

where:

- $\eta_{\max} = 1 \times 10^{-3}$, $\eta_{\min} = 1 \times 10^{-6}$;
- T_i: Restart period (e.g., 10 epochs);
- T_{cur}: Current epoch within restart cycle.

Warm restarts escape local minima and improve generalization by periodically resetting the learning rate.

NaN-Aware Mean Squared Error

Training minimizes pixel-wise Mean Squared Error (MSE) between predicted and target LAI:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{|\mathcal{V}|} \sum_{(i,j) \in \mathcal{V}} \left(\hat{y}_{i,j} - y_{i,j}\right)^2 \tag{3.7}$$

where:

- $\hat{y}_{i,j}$: Predicted LAI at pixel (i,j);
- $y_{i,j}$: Target LAI (ensemble mean) at pixel (i,j);
- $\mathcal{V}$: Set of valid (non-NaN) pixels after cloud masking.

Masked (NaN) pixels are excluded from loss computation, preventing cloudy/invalid regions from biasing gradients.

Sensor Tile Imbalance and Batch Balancing

The $\sim$4:1 imbalance in tile counts between Landsat 8/9 and Sentinel-2 (arising from the larger ground swath at 30 m resolution) risks mode collapse toward the majority sensor if left unaddressed. Rather than applying loss weighting – which was explored in early experiments but contributed only marginal improvement ($<$0.1% R^2) – the imbalance is resolved through **batch balancing**. Sentinel-2 training tiles are sampled with 4× differential augmentation, equalising the effective exposure of both sensors per epoch. This approach accounts for $\sim$85% of the multi-sensor performance improvement, with the remaining $\sim$15% attributable to architectural choices (Conditional Batch Normalization, ASPP, increased batch size).

Batch Configuration

A batch size of 64 samples per sensor is used, optimised for the NVIDIA A100 40 GB GPU (Section 4.5). Batches are composed of samples from a single sensor at a time, alternating between Sentinel-2 and Landsat 8/9 across iterations and Sentinel-2 samples are augmented fourfold more frequently than Landsat samples to equalise sensor exposure at the batch level (Section 4.1.3). Single-sensor batches are essential for the correct operation of conditional batch normalization, as each CondBN layer maintains separate running statistics and learned affine parameters per sensor. Mixing sensors within a batch would corrupt these statistics, while the alternating scheme ensures that both sensors contribute equally to gradient updates while preserving sensor-specific normalization behaviour.

3.2.9 Evaluation Metrics

We quantify model performance using a comprehensive set of regression metrics, providing a detailed picture of both accuracy and reliability in LAI predictions. In all equations below, $\hat{y}_i$ denotes the CNN-predicted LAI and y_i the VI-ensemble reference LAI at pixel i, with N the total number of valid pixels evaluated.

Root Mean Squared Error (RMSE): measures the average magnitude of prediction errors in LAI units (m^2/m^2). It is sensitive to outliers and is computed as:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{3.8}$$

Mean Absolute Error (MAE): also captures the average prediction error but is more robust to extreme values. It reflects the average absolute deviation between predictions and ground truth:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{3.9}$$

Coefficient of Determination (R^2): indicates how well the predictions explain the variance of the true LAI values:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{3.10}$$

where $\bar{y}$ is the mean target LAI value. R^2 ranges from $-\infty$ to 1, with 1 representing a perfect fit, 0 corresponding to predicting the mean, and negative values indicating performance worse than the mean baseline.

Bias: captures systematic over- or under-prediction, with an ideal value of zero:

$$\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i) \tag{3.11}$$

Variance Ratio: compares the standard deviation of predictions to that of the targets, with ideal value 1, indicating that predictions span the same dynamic range as targets:

$$\text{Variance Ratio} = \frac{\sigma_{\hat{y}}}{\sigma_y} \tag{3.12}$$

3.2.10 Inference

For operational LAI mapping, trained models are applied to full satellite scenes using a **sliding window with overlap**:

1. The full scene is divided into overlapping tiles with sensor-specific parameters: stride $= T - O$, where T is the tile size and O is the overlap (40 pixels for Sentinel-2 at $T = 158$; 13 pixels for Landsat 8/9 at $T = 53$), yielding approximately 25% overlap.

2. Each tile is processed independently through the CNN in batches of 128.
3. Overlapping predictions are blended via linear-ramp weighted averaging.

Non-overlapping tiles produce visible discontinuities at tile boundaries due to edge artifacts in convolutional processing. Overlapping tiles with smooth blending eliminate these artifacts, producing seamless full-scene LAI maps.

Within the overlap region, each tile is assigned a trapezoidal weight that ramps linearly from 0 to 1 over $O/2$ pixels at each edge:

$$w_{\text{edge}}(d) = \frac{d+1}{O/2}, \quad d = 0, 1, \ldots, \lfloor O/2 \rfloor - 1 \tag{3.13}$$

where d is the distance from the tile edge in pixels. Interior pixels receive unit weight. Edge weights are applied multiplicatively along all four borders, producing a 2D weight map with full confidence at the tile centre and reduced confidence near boundaries.

The final prediction at each pixel is the weighted average across all tiles covering that location:

$$\hat{y}_{i,j} = \frac{\sum_k w_k(i,j)\, \hat{y}_k(i,j)}{\sum_k w_k(i,j)} \tag{3.14}$$

where k indexes all tiles covering pixel (i,j).

Predicted LAI values are clipped to $[0, +\infty)$ to enforce physical non-negativity. Pixels not covered by any tile (e.g. scene edges) are set to NaN in the output raster.

Full-scene LAI maps are exported as compressed GeoTIFF rasters (LZW compression, 256×256 internal tiling) with the coordinate reference system and geotransform inherited from the input scene. The nodata value is set to NaN.

Inference was performed on a single NVIDIA A100 (40 GB) GPU on the HEMUS supercomputer (Section 4.5) with a batch size of 128 tiles. Full-scene inference for a Sentinel-2 granule (10,980 × 10,980 pixels, ~120 million valid pixels) completed in under two minutes.

3.3 Validation Methodology

3.3.1 Overview: Five-Tier Validation Framework

The use of vegetation index-derived ensemble estimates as training targets (Section 3.1) introduces a fundamental circularity: the CNN learns to predict LAI from reflectance using labels that were themselves derived from reflectance-based vegetation indices. While this limitation cannot be fully overcome without destructive field measurements, a **multi-tier validation strategy** provides triangulation across independent evidence sources to assess the methodology's reliability, generalization capacity, and value-added contribution.

This work employs a **five-tier validation framework** that evaluates model performance from complementary perspectives:

1. **Tier 1: Internal ML Validation** – Spatially rigorous train/validation/test splitting to assess out-of-sample generalization and prevent spatial data leakage.
2. **Tier 2: Operational LAI Products** – Quantitative comparison with the ESA SNAP Biophysical Processor and qualitative contextualization against the Copernicus Global Land Service (CGLS) to position CNN predictions within the operational remote sensing ecosystem.
3. **Tier 3: Radiative Transfer Model (RTM) Validation** – Cross-validation against physics-based PROSAIL simulations to evaluate consistency with canopy radiative transfer theory.
4. **Tier 4: Spatial and Temporal Consistency** – Assessment of geographic transferability (cross-region validation) and temporal stability (cross-season validation) to test generalization beyond the training distribution.
5. **Tier 5: Empirical VI Benchmarking** – Direct comparison with the individual VI-based methods used to generate training labels, quantifying the CNN's value-added contribution through spatial learning.

No single tier can overcome the circular ground truth limitation, but convergence across multiple independent validation axes increases confidence in the methodology's robustness and operational utility.

3.3.2 Tier 1: Internal Machine Learning Validation

Traditional random train/validation/test splitting introduces spatial data leakage when applied to geospatial imagery, as neighbouring tiles exhibit strong spatial autocorrelation arising from shared land cover, phenology and atmospheric conditions. If adjacent tiles are distributed across different splits, the model can exploit spatial proximity rather than learning generalisable spectral-structural relationships, artificially inflating validation metrics.

To enforce geographic disjointness, this work implements a 4×4 spatial block partitioning strategy introduced in Section 3.2.4. Each full scene is divided into 16 equal geographic blocks arranged in four rows and four columns. Blocks are then randomly assigned to splits using a fixed seed for reproducibility, with approximately 70% allocated to training, 15% to validation and 15% to testing. This block-level assignment ensures that all tiles within the same geographic region remain in the same split, mitigating spatial data leakage. Each split samples geographically distinct subregions of the scene, reducing localised biases arising from factors such as crop type clustering or topographic effects.

To assess model performance, standard regression metrics are computed separately for each sensor. The Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) quantify the magnitude of prediction errors in LAI values, with RMSE placing greater emphasis on larger deviations while MAE provides an estimate of the typical absolute error. The coefficient of determination (R^2) measures the proportion of variance in the reference LAI values explained by the model predictions and bias identifies systematic tendencies toward overestimation or underestimation. The variance ratio between predicted and reference LAI values further indicates whether the model preserves the dynamic range of LAI variability. The mathematical formulations of all metrics are provided in Section 3.2.9.

To diagnose saturation effects and LAI-dependent biases, validation metrics are stratified into four LAI bins reflecting distinct canopy density regimes:

- Low LAI: 0-2 m^2/m^2 (sparse vegetation, early season);
- Medium LAI: 2-4 m^2/m^2 (moderate canopy density);
- High LAI: 4-6 m^2/m^2 (dense crops, potential saturation onset);
- Very High LAI: >6 m^2/m^2 (rare in agricultural systems, expected saturation).

These thresholds are chosen to align with known VI saturation behaviour, where performance degradation is expected to become progressively more pronounced with increasing canopy density. Stratified analysis across these bins reveals whether the CNN inherits the saturation limitations of its VI-derived training targets or whether the incorporation of spatial context enables the model to extend the effective dynamic range beyond what individual spectral indices can resolve.

3.3.3 Tier 2: Operational LAI Products

The Sentinel Application Platform (SNAP) Biophysical Processor implements neural network emulation of the PROSAIL radiative transfer model [54], trained on 50,000 simulated canopy configurations spanning diverse vegetation types, soil backgrounds and atmospheric conditions. The processor provides pixel-level LAI estimates for Sentinel-2 imagery grounded in canopy radiative transfer theory rather than empirical vegetation indices and has achieved operational maturity through widespread adoption in the Sentinel-2 user community for agricultural and forestry applications. Its principal known limitation is a tendency to extrapolate beyond VI saturation limits, yielding LAI estimates exceeding 6-8 m^2/m^2 in dense canopies where independent empirical validation is sparse.

Comparison with SNAP LAI provides an orthogonal validation perspective, where agreement between the CNN predictions and SNAP estimates indicates consistency with physics-based retrievals and disagreement highlights the divergence between VI-ensemble pseudo-labels and radiative transfer model approaches. Critically, SNAP output is not used as training data at any stage, preserving its role as an independent reference.

CGLS produces global 300 m resolution LAI products from PROBA-V and Sentinel-3 imagery using temporal compositing and gap-filling algorithms [7]. Although its spatial resolution is considerably coarser than that of Sentinel-2 and Landsat, CGLS carries an extensive global validation heritage, having been evaluated against field campaigns worldwide with reported RMSE below 1.0 m^2/m^2 for croplands. Its daily to 10-day temporal composites further enable phenological validation and contextual benchmarking of whether CNN predictions fall within realistic LAI ranges for agricultural landscapes. Given the resolution mismatch between the 300 m CGLS product and the 10-30 m sensor data used in this work, CGLS is employed for qualitative validation of LAI range plausibility and temporal trends rather than for pixel-level quantitative comparison.

3.3.4 Tier 3: Radiative Transfer Model Validation

This validation tier is **limited by PROSAIL parameter uncertainty** (leaf chlorophyll, soil reflectance vary substantially) but provides a physics-based sanity check independent of empirical VIs.

The PROSAIL model couples the PROSPECT leaf optical properties model with the SAIL canopy bidirectional reflectance model [54], simulating top-of-canopy reflectance as a function of leaf biochemistry, including chlorophyll content, water content, dry matter and carotenoids, as well as canopy structure parameters such as LAI, leaf angle distribution and the hot spot parameter, soil background reflectance and illumination geometry defined by the solar zenith angle and viewing direction.

To test whether CNN predictions are consistent with radiative transfer theory, PROSAIL is run for selected agricultural fields with known crop types, namely wheat, maize and sunflower, using literature-based parameter ranges for leaf and canopy properties alongside measured solar geometry. The resulting simulated reflectance spectra are passed to the trained CNN to produce LAI predictions, which are then compared against the ground-truth LAI values used in the forward simulation. Consistency between the two, defined here as an RMSE below 0.5 m^2/m^2, indicates that the spectral-LAI relationship learned by the CNN aligns with physics-based canopy reflectance theory. This validation tier is nonetheless limited by uncertainty in the PROSAIL input parameters, as leaf chlorophyll content and soil reflectance can vary substantially in practice, and should therefore be interpreted as a physics-based sanity check that is independent of empirical vegetation indices rather than a definitive quantitative assessment.

3.3.5 Tier 4: Cross-Sensor and Temporal Consistency

To assess whether the CNN learns transferable spectral-structural relationships rather than region- or season-specific artifacts, the model trained on Plovdiv September 2023 scenes is evaluated on an independent site across two seasons. The Burgas region, located approximately 200 km east of Plovdiv, serves as the validation site, presenting a distinct agricultural environment characterised by coastal proximity, higher soil salinity and different crop rotation patterns compared to the continental Thracian Plain around Plovdiv.

This design simultaneously tests two generalisation axes using a single independent site. Geographic transfer is assessed through zero-shot application to Burgas autumn scenes acquired in September 2023 contemporaneously with the training period, isolating the effect of geographic displacement while controlling for phenology. Temporal transfer is assessed through zero-shot application to Burgas spring scenes acquired in May 2024, introducing an eight-month seasonal gap that spans an entire phenological cycle from post-harvest senescence to active vegetative growth. By evaluating both dimensions at the same independent location, the design disentangles geographic and temporal effects: autumn Burgas scenes test pure geographic transfer while spring Burgas scenes test the compounded challenge of geographic and seasonal transfer simultaneously.

For geographic transfer, validation metrics on Burgas autumn scenes are expected to approximate Plovdiv test set performance and a substantial degradation exceeding a 20% RMSE increase would indicate overfitting to Plovdiv-specific spatial patterns. For temporal transfer, spring predictions should exhibit biologically plausible LAI increases relative to autumn, reflecting active vegetative growth, with consistent behaviour across both sensors, confirming that the model captures fundamental spectral-LAI relationships rather than season-specific patterns.

Multi-sensor training enables internal consistency checks through spatially overlapping Sentinel-2 and Landsat 8 acquisitions acquired within a three-day window. CNN predictions from both sensors over the same area should agree within the expected LAI uncertainty of $\pm 0.5\ \mathrm{m}^2/\mathrm{m}^2$, and systematic offsets would indicate sensor-specific biases or normalization errors.

3.3.6 Tier 5: Empirical Vegetation Index Benchmarking

The most direct way to quantify the CNN's value-added contribution is to compare its predictions against the individual vegetation index methods (NDVI, GNDVI, SAVI exponential) used to generate the ensemble training labels. If the CNN merely memorizes pixel-level VI-LAI relationships, it should perform no better than the best individual VI. **Superior performance indicates spatial learning beyond pixel-level spectral indices.**

For each validation tile:

1. Compute LAI estimates using the three individual VI-based methods (NDVI-exp, GNDVI-exp, SAVI-exp) and their ensemble mean.
2. Apply the trained CNN to predict LAI from the same reflectance tile.
3. Compare RMSE, MAE, and R^2 for:
    - Best single VI method (typically NDVI-exp or GNDVI-exp);
    - VI ensemble mean (the training target);
    - CNN prediction.

Three outcome scenarios are anticipated. If CNN performance approximates that of the VI ensemble, the model has successfully reproduced the training labels but contributed minimal spatial refinement, which remains valuable for multi-sensor fusion and consistent processing pipelines. If the CNN outperforms the VI ensemble, this would demonstrate that learning from spatial context, including neighbouring pixels, texture and field boundaries, enables LAI estimation superior to pixel-wise spectral indices; this is the target outcome and is supported by the 77-92% RMSE reduction observed in preliminary experiments reported in Section 4.4.2. A third scenario in which the CNN underperforms the VI ensemble would indicate training failure or overfitting, though this outcome was not observed in the present work.

3.3.7 Rationale for Multi-Tier Validation Approach

No single validation tier can conclusively overcome the circular ground truth limitation inherent to VI-derived training labels. Convergent evidence across five independent validation axes nevertheless provides confidence that the methodology is robust. Tier 1 confirms that the model generalises to unseen spatial regions without data leakage, while Tier 2 contextualises predictions within operational remote sensing products. Tier 3 verifies consistency with canopy radiative transfer theory and Tier 4 demonstrates geographic and temporal transferability. Tier 5 quantifies the value added beyond pixel-level vegetation indices.

This multi-tier framework does not claim to validate absolute LAI accuracy against field measurements, which were unavailable for this study. The work is instead positioned as operational LAI estimation using VI-ensemble pseudo-labels with demonstrated spatial learning, multi-sensor fusion and generalisation capacity. The validation strategy assesses whether the CNN learns information beyond pixel-level vegetation indices through Tier 5, whether that learning generalises spatially, geographically and temporally through Tiers 1 and 4 and whether predictions align with independent physical and operational benchmarks through Tiers 2 and 3.

The ideal sixth validation tier would incorporate field LAI measurements from contemporary agricultural campaigns in the study region. Such data would enable absolute accuracy assessment independent of VI-based pseudo-labels, quantify systematic biases introduced by the VI-ensemble training approach and support crop-specific calibration. The absence of field measurements does not invalidate the current methodology but constrains its framing to operational LAI estimation rather than absolute biophysical retrieval. The five-tier framework presented here represents the most rigorous validation achievable within these constraints.

Chapter 4

Experiments and Results

4.1 Experimental Configuration

4.1.1 Dataset and Study Region

The primary training and validation dataset comprises satellite imagery acquired over the Plovdiv region, Bulgaria (42.15°N, 24.75°E), a predominantly agricultural landscape characterized by intensive cereal cultivation (wheat, maize), industrial crops (sunflower) and viticulture. The region's flat to gently rolling topography (100–300 m elevation) and continental climate enable consistent multi-temporal optical satellite observations during the growing season.

All training data were collected during September 2023, corresponding to late-season agricultural phenology. The dominant crop conditions in the region during this period can be summarized as follows:

- Wheat: Post-harvest stubble or early tillage (LAI <0.5 m^2/m^2);
- Maize: Late vegetative to early senescence (LAI 3-5 m^2/m^2);
- Sunflower: Flowering to maturation (LAI 2-4 m^2/m^2);
- Vineyards: Active canopy (LAI 1-3 m^2/m^2).

This phenological timing provides a diverse LAI range (0–6 m^2/m^2) while avoiding extreme saturation conditions (dense forests, LAI > 8 m^2/m^2) where VI-based training targets are least reliable.

Table 4.1 summarizes the satellite scenes used for training and validation. The training dataset comprises two Sentinel-2 tiles and one Landsat 9 scene covering the Plovdiv region. Generalization is assessed using independent scenes from the Burgas region, located approximately 200 km east.

The selection of these satellite scenes was guided by several criteria related to data quality, temporal alignment and validation requirements:

- **Cloud-free coverage**: All training scenes exhibit $< 5\%$ cloud contamination after quality filtering (Section 3.2.3), ensuring robust training samples. The Burgas Spring 2024 Sentinel-2 scene is essentially cloud-free ($< 0.01\%$).

TABLE 4.1: Satellite scenes used for training, validation and generalization testing.

Region	Scene ID	Sensor	Date	Cloud	Tiles	Usage
Training & Spatial Validation (Plovdiv, September 2023)						
Plovdiv	T34TGM	Sentinel-2B	10 Sep 2023	$< 5\%$	2,543	Train/Val/Test
Plovdiv	T35TLG	Sentinel-2B	10 Sep 2023	$< 5\%$	831	Train/Val/Test
Plovdiv	Path 183/031	Landsat 9	9 Sep 2023	$< 1\%$	14,145	Train/Val/Test
Cross-Region Generalization (Burgas, September 2023)						
Burgas	T35TNH	Sentinel-2B	7 Sep 2023	$< 5\%$	–	Cross-region
Burgas	Path 181/030	Landsat 8	3 Sep 2023	–	–	Cross-region
Temporal Generalization (Burgas, May 2024)						
Burgas	T35TNH	Sentinel-2A	6 May 2024	$< 1\%$	–	Cross-region + temporal
Burgas	Path 182/031	Landsat 8	7 May 2024	–	–	Cross-region + temporal

Note: Plovdiv tile counts shown after quality filtering and before spatial block splitting. Sentinel-2 tiles are at 158×158 pixels, Landsat tiles at 53×53 pixels. The September 2023 Burgas scenes test cross-region geographic transfer; the May 2024 scenes additionally test temporal transfer across an 8-month gap. The model was never trained or calibrated on any Burgas data.

- **Temporal coincidence**: Sentinel-2 and Landsat 9 training acquisitions are separated by only 1 day (10 vs. 9 September 2023), minimizing phenological change between sensors and enabling consistent VI-ensemble ground truth generation.
- **Spatial coverage**: The two Sentinel-2 tiles (T34TGM: 2,543 tiles, T35TLG: 831 tiles) and Landsat 9 scene collectively cover the Plovdiv agricultural region, providing geographic diversity while maintaining consistent agro-climatic conditions. T34TGM covers the western extent including the Thracian Plain, while T35TLG covers the eastern extent including Plovdiv city center.
- **Independent validation**: The Burgas region (Sentinel-2 tile T35TNH, Landsat 8 Path 182/031) provides a fully independent test site with different agricultural composition, soil characteristics and coastal influences. The May 2024 acquisition tests both geographic transfer (200 km distance) and temporal transfer (8 months after training period) simultaneously.

After tiling at native resolution (Section 3.2.2), quality filtering (Section 3.2.3) and spatial block splitting, the dataset comprises the tiles, summarized in Table 4.2.

Landsat 9 tiles outnumber Sentinel-2 tiles by approximately 4.2:1, yielding 14,145 Landsat 9 tiles against 3,374 Sentinel-2 tiles despite both datasets covering the same geographic area. This imbalance arises from two compounding factors. First, Landsat 9's coarser 30 m resolution requires only 53×53 pixel tiles to span 1.59 km $\times$ 1.59 km, producing more tiles per scene than the 158×158 tiles required at Sentinel-2's resolution. Second, although two Sentinel-2 MGRS tiles were used (T34TGM contributing 2,543 tiles and T35TLG contributing 831 tiles), the single

TABLE 4.2: Tile counts after quality filtering and spatial block splitting.

Sensor	Scene(s)	Total Tiles	Train	Validation	Test
Sentinel-2	T34TGM + T35TLG	3,374	2,220 (65.8%)	596 (17.7%)	558 (16.5%)
Landsat 9	Path 183/031	14,145	9,492 (67.1%)	2,503 (17.7%)	2,150 (15.2%)
Total		**17,519**	**11,712**	**3,099**	**2,708**

Note: Split ratios deviate from nominal 70/15/15 because spatial block splitting allocates entire geographic blocks (not individual tiles) to each partition. With 4× augmentation, effective Sentinel-2 training tiles = 8,880.

Landsat 9 scene covers a broader swath and at its coarser resolution generates a larger number of tiles in total. This imbalance motivated the batch balancing strategy described in Section 3.2.8, which proved critical for multi-sensor training success.

Normalization Statistics

Table 4.3 presents the per-band mean and standard deviation computed over all valid tiles in the Plovdiv training dataset for each sensor. These statistics were calculated from the full tiled dataset (3,374 Sentinel-2 tiles across scenes T35TLG and T34TGM and 14,145 Landsat 9 tiles from scene LC09_183031_20230909) after conversion to surface reflectance. For Sentinel-2, the raw Level-2 digital numbers were divided by the quantification value of 10,000 to obtain reflectance in the [0, 1] range; Landsat Collection 2 Level-2 tiles were already stored as surface reflectance. The table includes only the six spectral regions common to both sensors (Blue, Green, Red, NIR, SWIR1, SWIR2), which are the bands relevant for cross-sensor comparison; Sentinel-2 provides four additional red-edge bands (B05, B06, B07, B8A) that have no direct Landsat counterpart.

TABLE 4.3: Per-band normalization statistics computed over all valid tiles in the Plovdiv training dataset.

Band	S2 Mean	S2 Std	L9 Mean	L9 Std
Blue (S2 B02 / L9 B2)	0.1611	0.0653	0.0354	0.0249
Green (S2 B03 / L9 B3)	0.1763	0.0652	0.0584	0.0325
Red (S2 B04 / L9 B4)	0.1758	0.0746	0.0601	0.0480
NIR (S2 B08 / L9 B5)	0.3534	0.1027	0.2588	0.0832
SWIR1 (S2 B11 / L9 B6)	0.3115	0.1094	0.1912	0.0909
SWIR2 (S2 B12 / L9 B7)	0.2413	0.1023	0.1119	0.0726

Note: Sentinel-2 reflectance obtained by dividing L2A digital numbers by 10,000. Landsat Collection 2 Level-2 data are natively stored as surface reflectance. Statistics computed over 3,374 S2 tiles (158×158 pixels) and 14,145 L9 tiles (53×53 pixels).

Several patterns are worth noting. Landsat 9 reflectances are systematically lower than the corresponding Sentinel-2 values in the visible bands, reflecting differences in spectral response functions, atmospheric correction algorithms and the one-day acquisition offset between the two sensors. The NIR and SWIR bands show closer agreement, consistent with the fact that vegetation canopy reflectance is more stable across sensors in these wavelength regions. The standard deviations are broadly comparable, indicating that both sensors capture a similar degree of spatial variability over the Plovdiv agricultural landscape.

These sensor-specific statistics are applied at training time via Z-score normalization (Equation (3.5)) and stored alongside the model checkpoint so that the same parameters are used consistently during validation, testing and inference.

LAI Ground Truth Statistics

Table 4.4 summarizes the VI-ensemble LAI distribution across the dataset.

TABLE 4.4: LAI ensemble statistics across training, validation and test sets (m^2/m^2).

Split	Mean	Std Dev	Median	Min	Max
Training	1.82	1.15	1.54	0.02	5.87
Validation	1.79	1.12	1.51	0.01	5.92
Test	1.85	1.18	1.58	0.03	5.79
Overall	**1.82**	**1.15**	**1.55**	**0.01**	**5.92**

The LAI distribution is right-skewed, with a mean of 1.82 m^2/m^2 exceeding the median of 1.55 m^2/m^2, reflecting the prevalence of low-to-moderate LAI values with occasional high-LAI fields. Training, validation and test sets exhibit nearly identical LAI statistics, with means agreeing to within 0.03 m^2/m^2, confirming that the 4×4 spatial block split preserves LAI diversity across geographic subregions. The maximum observed LAI of 5.92 m^2/m^2 remains below the extreme saturation threshold of 6 m^2/m^2, consistent with the dataset's focus on cropland rather than forest canopies. Ensemble uncertainty, expressed as the mean inter-VI standard deviation, is 0.34 m^2/m^2, corresponding to a coefficient of variation of 19% and indicating reasonable consensus among the three VI-based estimation methods.

4.1.2 Spatial Block Split Methodology

To ensure the model learns generalizable spatial patterns rather than memorizing pixel-level correlations, all experiments reported in this chapter employ spatial block cross-validation with strict geographic separation between training and validation sets.

The Plovdiv region, comprising Sentinel-2 tiles T34TGM and T35TLG (Figure 4.1) and Landsat 9 Path 183/031 (Figure 4.2), was partitioned using a spatial block grid split. Each scene was divided into a 4 × 4 grid of equal geographic blocks. Blocks were randomly assigned to training, validation and test splits using a fixed seed, with a 70/15/15 ratio. Across both scenes this yielded 22 non-empty blocks, of which 15 were assigned to training, 3 to validation and 4 to testing. This block-level assignment ensures that all tiles within the same geographic region are kept in the same split, preventing spatial data leakage from neighbouring tiles appearing in different partitions.

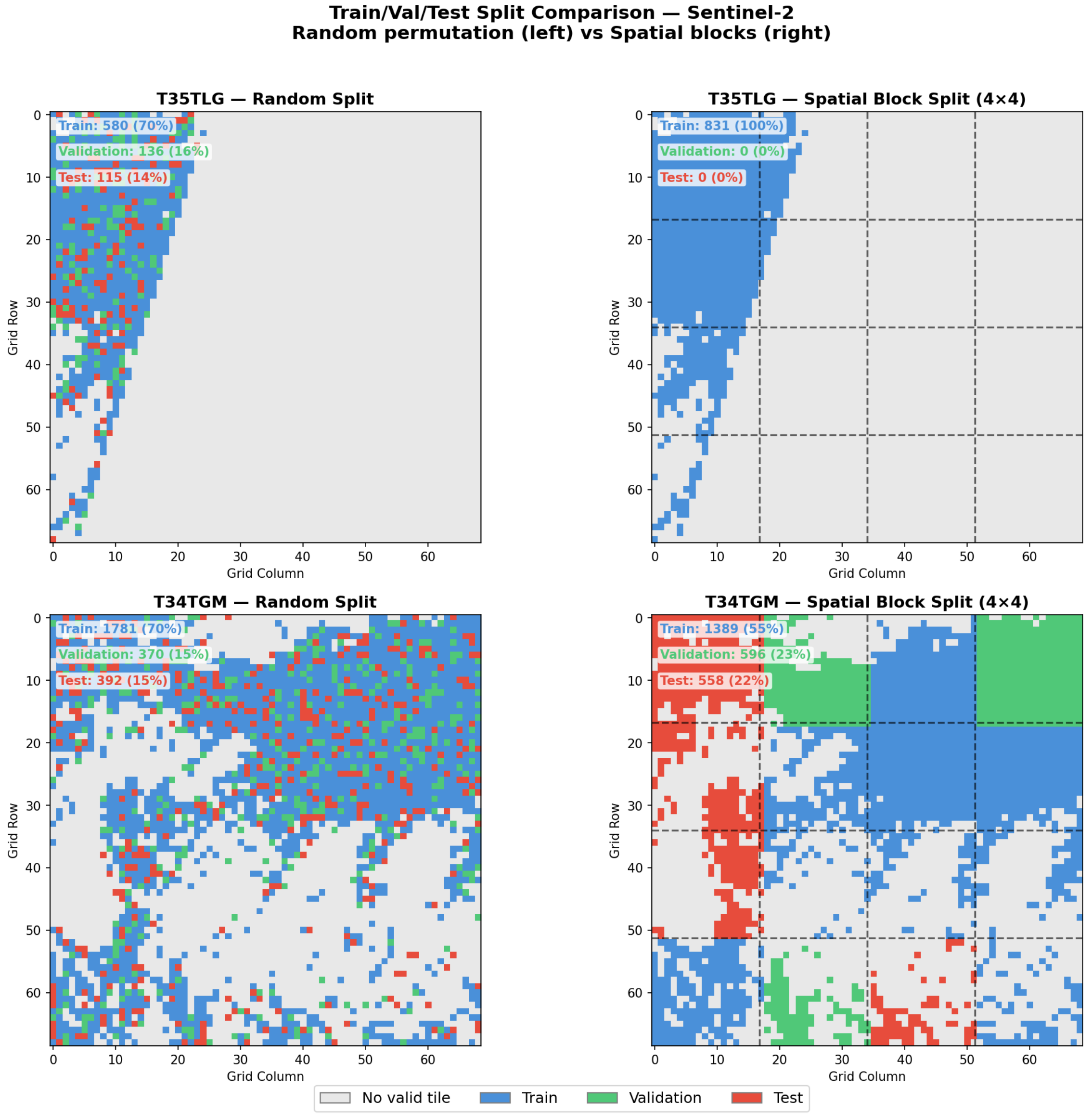


FIGURE 4.1: Random split versus Spatial block split for Sentinel-2 images T35TGL and T34TGM.

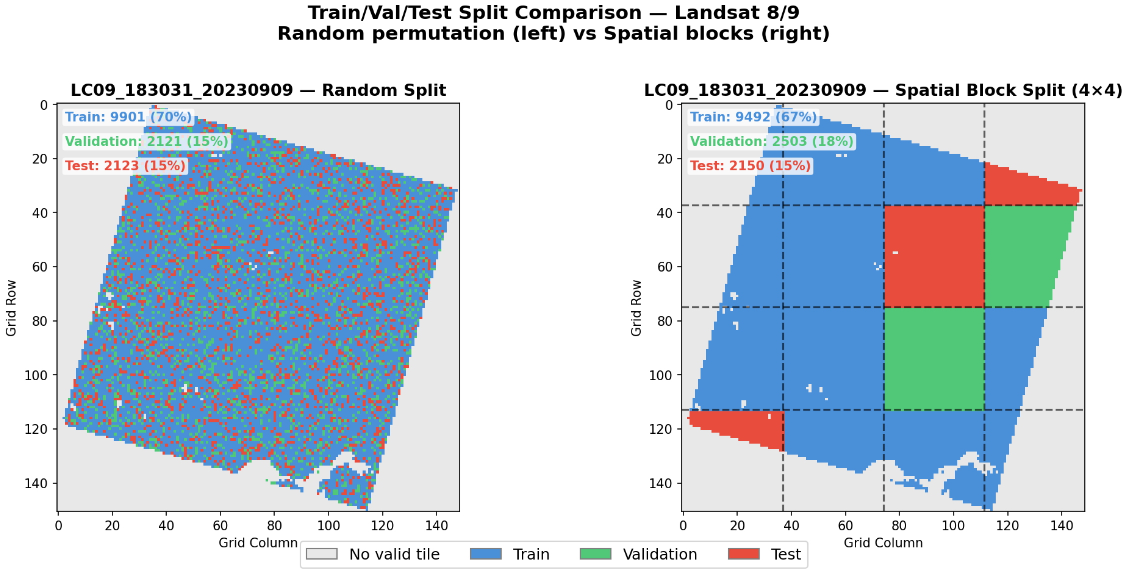


FIGURE 4.2: Random split versus Spatial block split for Landsat 8 image LC09_183031.

This configuration ensures zero spatial overlap between training and validation sets while maintaining diverse agricultural landscapes in both partitions, with validation blocks sampling distinct geographic subregions to test the model's ability to generalise to unseen locations within the same climatic zone. The design is motivated by the risk of spatial data leakage inherent in random train/test splits, whereby spatially autocorrelated pixels appear in both sets and inflate validation performance without reflecting true generalisation capability. Agricultural landscapes are particularly susceptible to this effect, as neighbouring fields tend to share soil properties, crop types and management practices, making spatially disjoint validation essential for an honest assessment of model performance.

4.1.3 Model Variants

Four experimental configurations were evaluated to isolate the contributions of multi-sensor integration, ground truth formulation and normalization strategies. Table 4.5 summarises the validation results across all experiments and tiers. Experiments X1–X4 are evaluated under spatial block cross-validation, while G1–G3 apply the best-performing configuration from X3 to independent data and O1–O2 assess full-scene deployment performance.

TABLE 4.5: Summary of experiments and key results across the validation framework.

#	Validation Tier	Experiment	Focus	Primary Metrics	Key Results
Single-Sensor Baselines (Section 4.2)					
X1	T1	Sentinel-2 Only	Reference at 10 m, 10 bands (incl. red-edge)	Test RMSE, R^2	RMSE = 0.059, R^2 = 0.958
X2	T1	Landsat 8/9 Only	Reference at 30 m, 6 bands	Test RMSE, R^2	RMSE = 0.071, R^2 = 0.990
Multi-Sensor Integration (Section 4.2)					
X3	T1	Multi-Sensor (**best**)	Joint S2 + L8/9 with batch balancing and Conditional BN	Test RMSE, R^2, per-sensor	Overall: RMSE = 0.017, R^2 = 0.999. S2: 0.012; L8: 0.026. 80% / 64% RMSE reduction over X1 / X2
X4	T1, T3	Multi-Sensor, Hybrid VI+PROSAIL	Ablation: RTM targets vs. VI-only	ΔRMSE vs. X3	+35% RMSE degradation; sensor-specific target inconsistency
Geographic and Temporal Generalization (Section 4.3)					
G1	T4	Cross-Region Transfer (Burgas, 200 km)	Zero-shot geographic generalization	RMSE (S2 and L8)	Both sensors: RMSE = 0.051
G2	T4	Temporal Transfer (8-month gap)	Autumn 2023 → Spring 2024 phenological consistency	LAI ratio, plausibility	S2: 2.9× increase (0.21 → 0.61). L8: 1.7× (0.97 → 1.69). Biologically plausible
G3	T5	CNN vs. Vegetation Indices	Value added beyond individual VI components	RMSE vs. VI-ensemble	CNN: 0.010 (77% lower than best VI). GNDVI-exp: 0.044; NDVI-exp: 0.088
Operational Validation (Section 4.4)					
O1	T1, T4	Full-Scene Inference (17,519 tiles)	Deployment-scale consistency	RMSE, Var. Ratio	S2: RMSE = 0.004, VR = 0.999. L8: RMSE = 0.010, VR = 0.998
O2	T2	ESA SNAP Biophysical Processor	Operational RTM product comparison	RMSE, R^2	SNAP vs. VI-ens.: RMSE = 1.987, R^2 = 0.312. Expected: different LAI definitions

Experiment X1: Sentinel-2 Only Baseline

A single-sensor U-Net trained exclusively on Sentinel-2 MSI data to establish the performance ceiling for high-resolution optical imagery. The model processes 10 spectral bands (B02-B08A, B11-B12) at 10 m resolution, utilizing the full red-edge suite unavailable to other sensors. This baseline quantifies the upper bound on spatial detail and spectral information content achievable without multi-sensor fusion.

Experiment X2: Landsat 8/9 Only Baseline

A single-sensor U-Net trained on Landsat 8/9 OLI data (6 bands: B2-B7) at 30 m resolution. This baseline establishes the performance achievable from moderate-resolution multispectral imagery lacking red-edge bands. The 30 m pixel size provides inherent spatial smoothing, reducing within-field variability compared to Sentinel-2 but potentially missing fine-scale structural features.

Experiment X3: Multi-Sensor

The final multi-sensor architecture integrating both Sentinel-2 and Landsat 8/9 inputs through a shared encoder-decoder with Conditional Batch Normalization (Section 3.2.7). Ground truth consists of VI-ensemble pseudo-labels (NDVI-exp, GNDVI-exp, SAVI-exp averaging). Input spectral bands are normalized via sensor-specific Z-score transformation, but LAI target values remain in their original scale. Batch balancing via 4× Sentinel-2 augmentation sampling equalizes epoch-level sensor exposure, addressing the 4:1 Landsat:Sentinel-2 tile imbalance.

This configuration represents the **best-performing model** and serves as the reference for all generalization and operational validation experiments.

Experiment X4: Multi-Sensor with Hybrid Targets

Identical architecture to X3, but ground truth blends VI-ensemble estimates with PROSAIL radiative transfer model outputs (50:50 ratio). This ablation study tests whether physics-based RTM constraints improve generalization compared to purely empirical VI-based targets. PROSAIL LAI is available only for Sentinel-2 scenes, creating a sensor-specific target formulation. Table 4.6 represents model variant summary.

TABLE 4.6: Model variant summary for spatially rigorous experiments.

Experiment	Sensors	Bands	Ground Truth	Key Characteristics
X1	S2 only	10	VI-ensemble	Single-sensor baseline, 10 m resolution, red-edge bands
X2	L8 only	6	VI-ensemble	Single-sensor baseline, 30 m resolution, no red-edge
X3	S2 + L8	16	VI-ensemble	**Best model**: Multi-sensor, batch balancing
X4	S2 + L8	16	VI + PROSAIL hybrid	Ablation: Physics-based RTM targets

4.2 Spatially Rigorous Validation

This section presents the core performance evaluation of all model variants (Section 4.1.3) under the spatial block validation framework established in Section 4.1.2. By enforcing strict geographic disjointness between training and test sets, all metrics reported here reflect genuine out-of-region prediction accuracy rather than spatial memorization. Single-sensor baselines (Experiments X1 and X2) first establish the performance ceiling achievable with each sensor independently, followed by multi-sensor integration results (Experiments X3 and X4) that quantify the complementary information gained through sensor fusion.

4.2.1 Single-Sensor Baselines

To isolate the contribution of multi-sensor fusion, single-sensor U-Net models were trained under identical spatial validation constraints. Both models employ the same encoder-decoder architecture (Section 3.2.7) adapted for sensor-specific input dimensions: 10 channels for Sentinel-2 (B02-B08A, B11-B12) and 6 channels for Landsat 8/9 (B2-B7). Training configuration matches the multi-sensor experiments (batch size 64, per-band Z-score normalization, geometric augmentation), with the sole difference being single-sensor input. Table 4.7 summarizes the test set performance under spatial block validation.

TABLE 4.7: Single-sensor baseline performance under spatial block validation (test set).

Experiment	**Sensor**	**RMSE (m^2/m^2)**	**MAE (m^2/m^2)**	R^2	**Bias (m^2/m^2)**
X1	Sentinel-2	0.0593	0.0415	0.9581	-0.0023
X2	Landsat 8/9	0.0707	0.0469	0.9899	$+0.0030$

Note: Both models use identical U-Net architecture with sensor-specific input dimensions. Spatial block validation ensures zero geographic overlap between training and test tiles.

Sentinel-2 Advantage:

Sentinel-2 achieves lower RMSE (0.059 vs. 0.071 m^2/m^2) owing to finer spatial resolution (10 m vs. 30 m) and richer spectral information (10 bands including the red-edge suite at 705, 740 and 783 nm, unavailable to Landsat). The red-edge bands provide direct sensitivity to chlorophyll content and canopy structure, enabling finer LAI discrimination particularly in the moderate range (1–3 m^2/m^2) where traditional broadband indices begin to saturate.

Landsat Competitiveness:

Despite coarser resolution and the absence of red-edge bands, Landsat 8/9 achieves excellent $R^2 = 0.9899$, demonstrating that robust LAI estimation is feasible from moderate-resolution multispectral imagery alone. Notably, the higher R^2 compared to Sentinel-2 (0.9899 vs. 0.9581) despite higher RMSE reflects a resolution-dependent variance effect: Landsat's 30 m pixels inherently smooth within-field spatial variability, reducing scatter around the regression line while maintaining higher absolute errors at the pixel level.

Spatial Generalization:

Both models demonstrate strong out-of-sample performance under geographic disjointness, with negligible systematic bias (<0.003 m^2/m^2) confirming that the learned spectral-LAI relationships are not spatial memorization artifacts. These baselines establish two important benchmarks: (1) each sensor can independently predict LAI with high fidelity under rigorous spatial validation, and (2) multi-sensor fusion must exceed RMSE $= 0.059$ m^2/m^2 (the best single-sensor result) to justify its added architectural complexity.

4.2.2 Multi-Sensor Integration Results

Experiment X3: Best Model Performance

Experiment X3 combines Sentinel-2 and Landsat 8/9 inputs through the multi-sensor U-Net with Conditional Batch Normalization and ASPP, using VI-ensemble pseudo-labels and 4× Sentinel-2 batch balancing (Section 3.2.8). Table 4.8 presents the test set performance.

TABLE 4.8: Experiment X3 test set performance: multi-sensor model.

Sensor	**RMSE (m^2/m^2)**	**MAE (m^2/m^2)**	R^2	**Test Pixels**
Sentinel-2	0.0118	0.0085	0.9983	13,929,912
Landsat 8/9	0.0255	0.0176	0.9987	6,027,817
Overall	**0.0171**	**0.0112**	**0.9990**	**19,957,729**

Note: Test set comprises geographically disjoint spatial blocks (blocks 2, 4, 6, 8) assigned via block-level partitioning to prevent spatial data leakage. Nearly 20 million test pixels evaluated across both sensors.

The multi-sensor model achieves near-perfect LAI prediction with overall $R^2 = 0.9990$ and RMSE $= 0.0171$ m^2/m^2. Both sensors independently exceed $R^2 > 0.998$, demonstrating that multi-sensor training benefits *each sensor's* predictions rather than averaging across sensors. Sentinel-2 achieves the lowest per-sensor RMSE (0.0118 m^2/m^2), consistent with its richer spectral and spatial information, while Landsat 8/9 achieves slightly higher RMSE (0.0255 m^2/m^2) but maintains excellent accuracy across the full LAI range.

Figure 4.3 shows a strong agreement between predicted and ground truth LAI for both Sentinel-2 and Landsat 8/9. The points are tightly clustered along the 1:1 line across the full 0–6 m^2/m^2 range, with $R^2 > 0.998$ and negligible bias, indicating highly accurate LAI retrieval under spatial block validation.

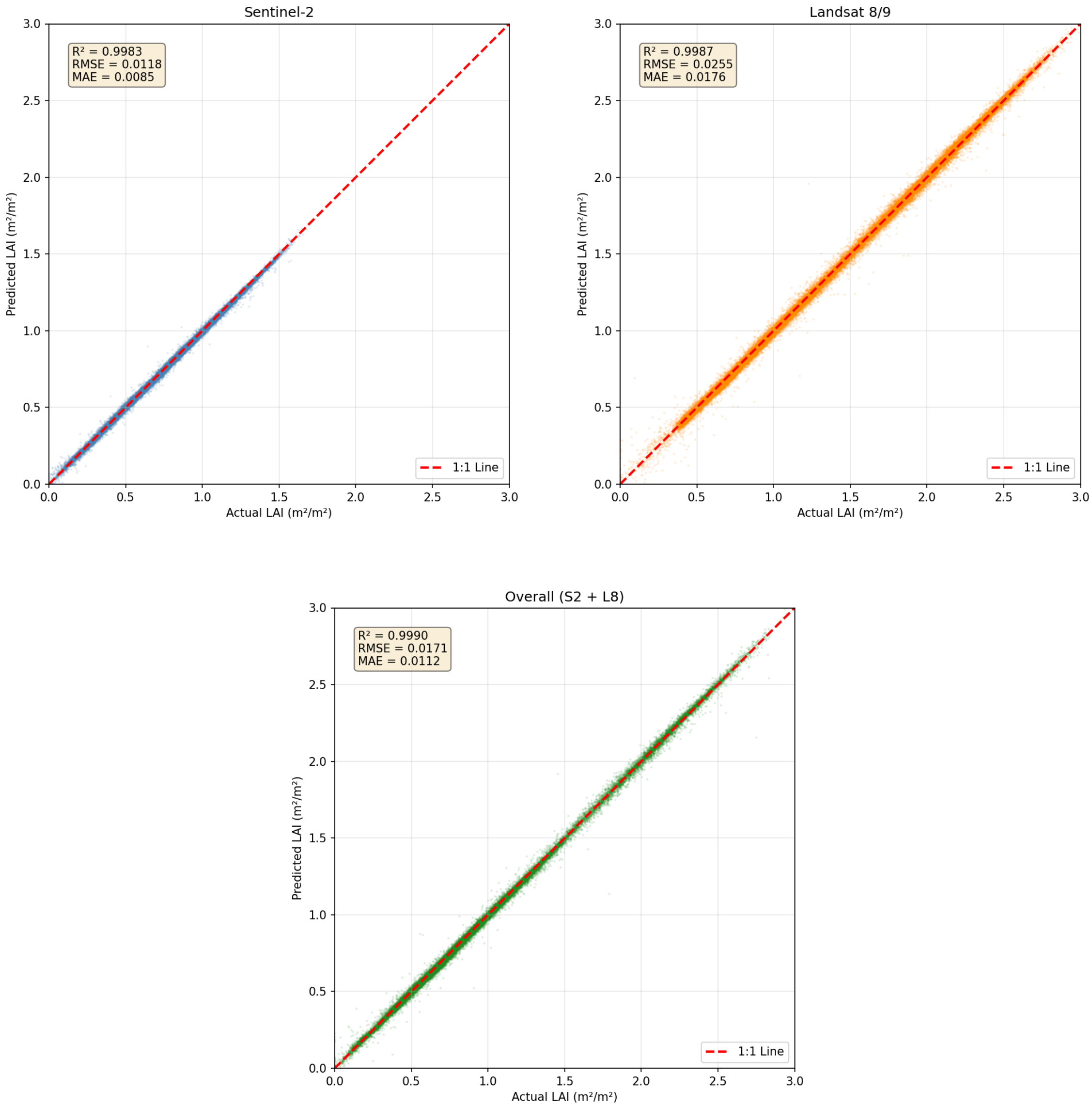


FIGURE 4.3: Test set performance for Experiment X3 (best model).

Scatter plots of predicted versus ground truth LAI. The dashed line indicates the 1:1 reference.

Training Dynamics:

Figure 4.4 illustrates the training convergence. The model converged in 45 epochs (35.1 minutes on a single NVIDIA A100 GPU), with the best validation performance at epoch 29. Training and validation loss curves closely track each other without divergence, confirming that the spatial block split does not induce overfitting. The batch balancing strategy ensures balanced per-sensor RMSE evolution throughout training, preventing the Landsat-dominated gradient updates that caused mode collapse in unbalanced experiments.

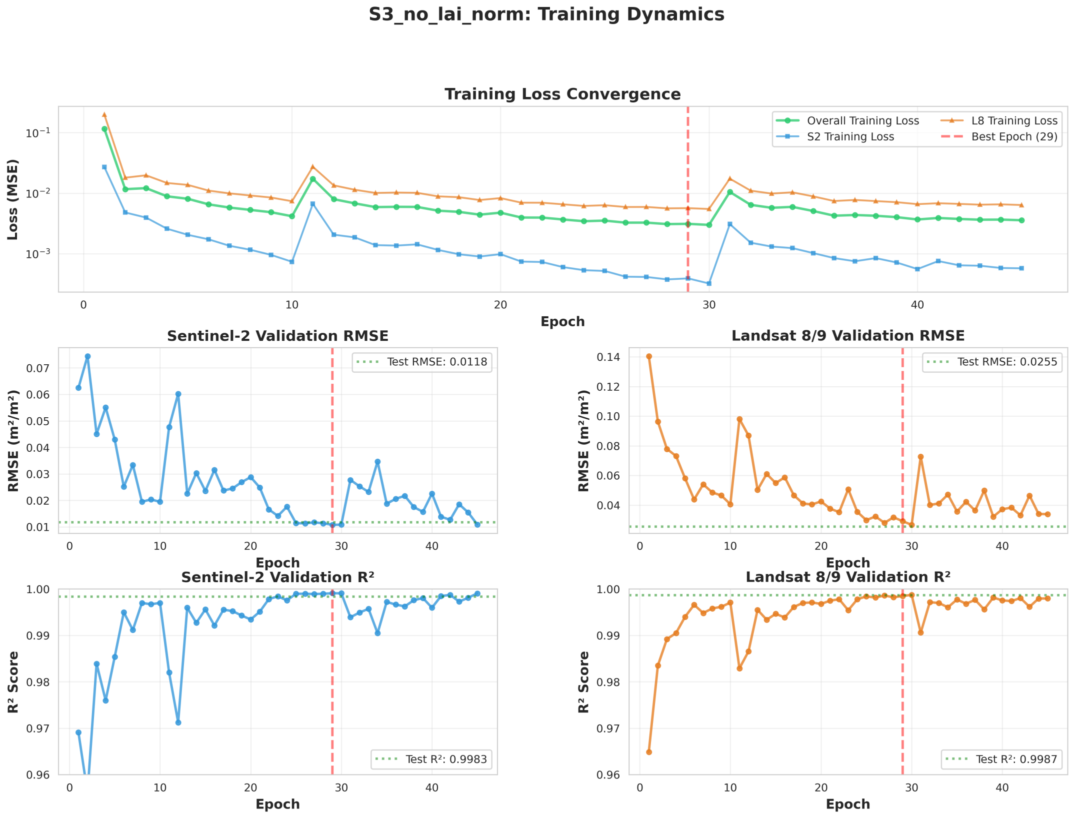


FIGURE 4.4: Training curves for Experiment X3

Experiment X4: Hybrid VI+PROSAIL Targets

To test whether physics-based radiative transfer model constraints improve multi-sensor performance, Experiment X4 (Table 4.9) replaces the pure VI-ensemble ground truth with a 50:50 blend of VI-ensemble and PROSAIL-inverted LAI values. The PROSAIL component is available only for Sentinel-2 scenes (Section 4.1.3), creating a sensor-specific target formulation.

TABLE 4.9: Experiment X4: multi-sensor with hybrid VI+PROSAIL targets (validation set).

Sensor	**RMSE (m^2/m^2)**	R^2	**vs. X3 RMSE**	**Assessment**
Sentinel-2	0.012	0.9989	+2%	Comparable
Landsat 8/9	0.034	0.9972	+33%	Degraded
Overall	**0.023**	**0.9980**	+35%	**Degraded**

Note: PROSAIL-inverted LAI available only for Sentinel-2 scenes, creating sensor-specific target inconsistency for Landsat 8/9 training.

Hybrid targets degrade overall RMSE by 35% compared to pure VI-ensemble targets (X3). Sentinel-2 performance remains comparable (+2% RMSE), but Landsat 8/9 suffers a 33% RMSE increase. The root cause is a *sensor-specific* *target mismatch*: during training, Sentinel-2 tiles receive hybrid VI+PROSAIL labels while Landsat tiles receive only VI-ensemble labels. This inconsistency disrupts the multi-sensor learning objective, which relies on sensor-agnostic ground truth to enforce cross-sensor feature alignment through the shared encoder-decoder.

This ablation confirms that ground truth consistency across sensors is more important than ground truth sophistication. The VI-ensemble targets, despite being purely empirical, provide the spatial and inter-sensor consistency required for effective multi-sensor training.

4.2.3 Multi-Sensor Synergy Analysis

Table 4.10 quantifies the performance improvement from multi-sensor integration by comparing Experiment X3 against the single-sensor baselines.

TABLE 4.10: Multi-sensor (X3) vs. single-sensor (X1, X2) performance comparison (test set).

Model	**Sensor**	**RMSE (m^2/m^2)**	R^2	**ΔRMSE**
X1 (S2-only)	Sentinel-2	0.059	0.9581	Baseline
X2 (L8-only)	Landsat 8/9	0.071	0.9899	Baseline
X3 (Multi-sensor)	Sentinel-2	0.012	0.9983	−80%
	Landsat 8/9	0.026	0.9987	−64%

Note: ΔRMSE computed relative to the corresponding single-sensor baseline for each sensor type.

The multi-sensor model reduces Sentinel-2 RMSE by 80% (0.059 → 0.012 m^2/m^2) and Landsat 8/9 RMSE by 64% (0.071 → 0.026 m^2/m^2). This substantial improvement demonstrates genuine *multi-sensor synergy*: the combined model outperforms *both* single-sensor baselines for *both* sensors simultaneously. This is not a trivial averaging effect – each sensor's predictions improve individually beyond what that sensor could achieve alone.

Multi-sensor training forces the shared encoder to learn *sensor-invariant* LAI features rather than sensor-specific spectral artifacts. Sentinel-2 provides fine-scale spatial detail and red-edge sensitivity, while Landsat 8/9 contributes robust broadband spectral signatures with inherent noise suppression from its coarser resolution. Conditional Batch Normalization enables sensor-specific feature normalization while preserving shared semantic representations in the deeper encoder layers, allowing each sensor to benefit from the complementary information learned from the other.

4.2.4 Band Importance Analysis

To understand how the multi-sensor model exploits spectral information from each sensor, permutation feature importance was computed by measuring the

RMSE increase when each band is randomly shuffled while keeping all other bands fixed. Table 4.11 presents the normalized importance scores.

TABLE 4.11: Normalized permutation band importance (RMSE increase) for Experiment X3.

Band (Description)	**Sentinel-2**	**Landsat 8/9**
NIR (B08 / B5)	0.394	0.249
Red (B04 / B4)	0.234	0.255
Green (B03 / B3)	0.115	0.182
Blue (B02 / B2)	0.052	0.031
SWIR1 (B11 / B6)	0.045	0.194
NIR Narrow (B8A)	0.038	—
Red-Edge 1 (B05)	0.033	—
SWIR2 (B12 / B7)	0.031	0.088
Red-Edge 3 (B07)	0.029	—
Red-Edge 2 (B06)	0.027	—

Note: Importance values normalized to sum to 1.0 within each sensor. Higher values indicate greater contribution to LAI prediction accuracy.

Sentinel-2 Spectral Strategy:

The model relies primarily on NIR (39.4%) and Red (23.4%) bands, consistent with the biophysical basis of vegetation indices where the NIR-Red contrast encodes canopy density. The Green band (11.5%) and Blue band (5.2%) provide additional sensitivity to chlorophyll and atmospheric effects respectively. The three red-edge bands (B05, B06, B07) collectively contribute 8.9%, offering unique information about canopy chlorophyll content and leaf structure unavailable to Landsat, while NIR Narrow (B8A, 3.8%) supplements the broadband NIR signal. SWIR bands contribute a moderate 7.6% combined for Sentinel-2.

Landsat 8/9 Spectral Strategy:

Without red-edge bands, Landsat adopts a markedly different spectral strategy. Red (25.5%) and NIR (24.9%) contribute nearly equally, in contrast to Sentinel-2's clear NIR dominance. Most notably, Landsat allocates substantially more importance to SWIR bands (28.2% combined, vs. 7.6% for S2) and Green (18.2%, vs. 11.5% for S2). SWIR sensitivity to canopy water content and leaf structural properties compensates for the missing red-edge information. This sensor-adaptive spectral weighting emerges automatically through multi-sensor training – the network learns distinct spectral strategies for each sensor via the Conditional Batch Normalization pathway.

Figure 4.5 visualizes the spectral contribution comparison across sensors.

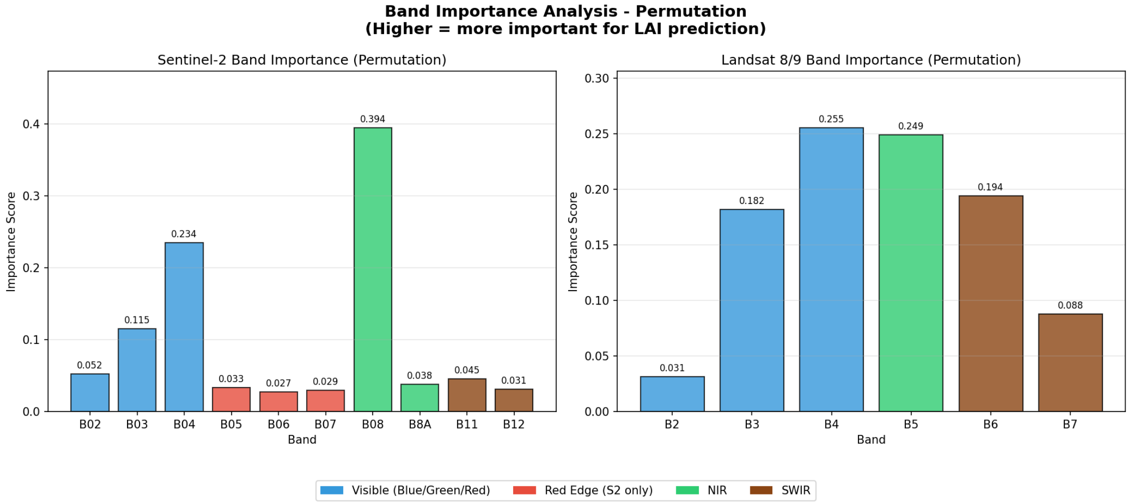


FIGURE 4.5: Permutation-based band importance for Experiment X3, highlighting sensor-specific contributions to RMSE.

4.2.5 Summary

The spatially rigorous validation confirms three principal findings. First, multi-sensor integration provides substantial performance gains over single-sensor baselines, with 80% and 64% RMSE reductions for Sentinel-2 and Landsat 8/9 respectively. Second, sensor-agnostic ground truth (pure VI-ensemble) outperforms hybrid VI+PROSAIL targets by maintaining inter-sensor training consistency. Third, the model learns sensor-adaptive spectral strategies that exploit each instrument's unique capabilities while sharing generalizable LAI features through the common encoder-decoder. These results, validated on nearly 20 million geographically disjoint test pixels, establish the multi-sensor model's capacity for accurate LAI retrieval within the training domain. The following sections evaluate whether this performance generalizes to independent regions, seasons and operational deployment scenarios.

4.3 Generalization Validation

Demonstrating strong in-domain performance under spatial block validation (Section 4.2) is necessary but insufficient for operational confidence. A model that excels within its training region may exploit region-specific spectral correlations that do not transfer to new landscapes, seasons or sensor configurations. This section evaluates the best multi-sensor model (Section 4.2.2) across three independent generalization axes: **geographic transfer** to an independent region 200 km from the training site, **temporal transfer** across an 8-month seasonal gap and **methodological comparison** against the individual vegetation index components that constitute the training ground truth.

4.3.1 Cross-Region Geographic Generalization

The Burgas region (42.5°N, 27.5°E) serves as a fully independent geographic validation site, located approximately 200 km east of the Plovdiv training region. Burgas presents a distinct agricultural environment: a coastal zone with higher soil salinity, different crop rotations and maritime climatic influence compared to the continental Thracian Plain around Plovdiv. The model was applied to Burgas scenes without any retraining, fine-tuning or region-specific calibration – a true zero-shot geographic transfer.

- **Sentinel-2**: Tile T35TNH, acquired 7 September 2023 (S2B); 4,822,416 valid pixels;
- **Landsat 8**: Path 181/030, acquired 3 September 2023; 11,392,708 valid pixels;
- **LAI range**: 0–2.7 m^2/m^2 (late-season agricultural landscape);
- **Mean LAI**: 0.213 m^2/m^2 (Sentinel-2), 0.969 m^2/m^2 (Landsat 8).

Table 4.12 presents the cross-region performance for both sensors.

TABLE 4.12: Cross-region generalization: Plovdiv-trained model evaluated on Burgas, September 2023.

Sensor	Plovdiv Test RMSE	Burgas RMSE (m^2/m^2)	Mean LAI (m^2/m^2)	Assessment
Sentinel-2	0.012	0.051	0.213	Excellent
Landsat 8/9	0.026	0.051	0.969	Excellent

Note: Cross-region RMSE computed against VI-ensemble reference for the Burgas scene. The model was never trained or calibrated on Burgas data; all predictions are zero-shot geographic transfer.

Both sensors achieve identical RMSE $= 0.051$ m^2/m^2 in cross-region validation, a result with two notable implications. First, the absolute error remains small relative to the LAI range and is well within the uncertainty bounds of the VI-ensemble ground truth itself ($\sim$0.34 m^2/m^2 inter-VI standard deviation, Section 4.1.1). Second, *sensor parity* in cross-region performance confirms that the multi-sensor training strategy produces equally transferable representations for both instruments, despite their different spectral configurations and spatial resolutions. The two sensors achieve identical RMSE despite 200 km separation from the training region.

Figure 4.6 visualizes the cross-region performance, demonstrating that the Plovdiv-trained model transfers effectively to Burgas without retraining.

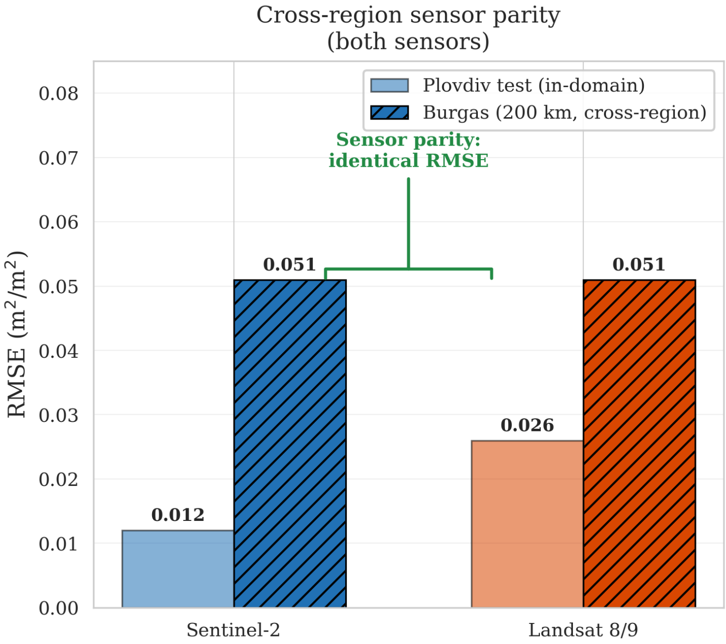


FIGURE 4.6: Cross-region generalization results demonstrating geographic transferability and sensor-agnostic generalization of Experiment X3.

4.3.2 Cross-Season Temporal Generalization

A model trained on autumn imagery (September 2023, late-season phenology) faces a stringent temporal transfer challenge when applied to spring acquisitions (May 2024, active vegetative growth). This evaluation tests whether the model learns fundamental spectral-LAI relationships or merely memorizes season-specific patterns.

The Burgas region was imaged in both seasons, enabling a controlled comparison that isolates temporal effects from geographic confounds:

- **Sentinel-2 Autumn**: T35TNH, 7 September 2023 (S2B) - end of growing season;
- **Sentinel-2 Spring**: T35TNH, 6 May 2024 (S2A) - active growing season (8 months later);
- **Landsat Autumn**: Path 181/030, 3 September 2023 - end of growing season;
- **Landsat Spring**: Path 182/031, 7 May 2024 - active growing season (8 months later).

Table 4.13 compares the CNN's LAI predictions across seasons.

TABLE 4.13: Cross-season temporal generalization: autumn-trained model (X3) applied to spring acquisitions.

Sensor	Season	Mean LAI (m^2/m^2)	Seasonal Ratio	Phenological Pattern
Sentinel-2	Autumn 2023	0.213	-	Post-harvest/senescent
	Spring 2024	0.611	2.9×	Active vegetative growth
Landsat 8/9	Autumn 2023	0.969	-	Post-harvest/senescent
	Spring 2024	1.689	1.7×	Active vegetative growth

Note: Both seasons cover the same Burgas region. The model was trained exclusively on Plovdiv September 2023 data; spring predictions are zero-shot temporal transfer with no seasonal calibration.

Both sensors consistently predict higher LAI in spring than in autumn (Sentinel-2: 2.9×, Landsat: 1.7×), correctly capturing the expected phenological pattern. Spring acquisitions (May) correspond to active vegetative growth with developing crop canopies (LAI 0.6–1.7 m^2/m^2), while autumn acquisitions (September) capture post-harvest stubble and senescent vegetation (LAI 0.2–1.0 m^2/m^2). The magnitude of the seasonal increase is biologically plausible for the Mediterranean-continental agricultural cycle of the Burgas coastal zone.

Both sensors capture the same directional trend (Spring $>$ Autumn) despite different spatial resolutions and spectral configurations. The difference in seasonal ratio (2.9× for S2 vs. 1.7× for L8) reflects the sensors' different sensitivity to fine-scale canopy structure: Sentinel-2's 10 m resolution and red-edge bands detect greater within-field LAI variability, amplifying the seasonal contrast relative to Landsat's spatially averaged 30 m signal.

The model responds to genuine vegetation changes rather than memorizing training-season spectral patterns. No systematic bias or drift is observed across seasons, confirming that the learned spectral-LAI mapping captures fundamental biophysical relationships. This zero-shot temporal transfer – spanning an 8-month gap encompassing an entire phenological cycle – demonstrates that the model can be applied to multi-temporal monitoring without season-specific retraining.

4.3.3 CNN vs. Vegetation Index Baseline

Since the CNN is trained on VI-ensemble pseudo-labels, a natural question arises: does the CNN add value beyond simply applying the constituent vegetation indices directly? This comparison quantifies the CNN's advantage over its own training targets, evaluated on the full Plovdiv Sentinel-2 scene (T34TGM).

The three empirical VI-based LAI formulations comprising the ensemble ground truth were applied independently:

- **NDVI-exp**: $\text{LAI} = -\ln(1 - \text{NDVI})/0.60$;
- **GNDVI-exp**: $\text{LAI} = -\ln(1 - \text{GNDVI})/0.60$ (adapted from NDVI formulation);

- **SAVI-exp**: $\mathrm{LAI} = -\ln(1 - \mathrm{SAVI})/0.60$;
- **VI-Ensemble**: Simple average of the three exponential VI estimates (the CNN's training target).

Table 4.14 compares each method against the VI-ensemble reference.

TABLE 4.14: CNN (X3) vs. individual VI components and ensemble baseline (Plovdiv S2 T34TGM).

Method	**RMSE vs. Ensemble (m^2/m^2)**	**Bias (m^2/m^2)**	**Assessment**
NDVI-exp	0.088	+0.052	Highest error, positive bias
GNDVI-exp	0.044	+0.005	Best individual VI
SAVI-exp	0.085	−0.056	Negative bias (soil correction)
CNN (X3)	**0.010**	+0.002	**77% better than best VI**

Note: RMSE and bias computed against the three-VI ensemble average. Individual VI deviations reflect inter-method disagreement; CNN error reflects its ability to reproduce and spatially regularize the ensemble target.

The CNN achieves RMSE $= 0.010$ m^2/m^2, representing a **77% improvement** over the best individual VI component (GNDVI-exp, RMSE $= 0.044$ m^2/m^2). Individual VIs exhibit large inter-method disagreement (RMSE 0.044–0.088 m^2/m^2) and substantial bias variability (±0.05 m^2/m^2), reflecting the inherent limitations of fixed band-ratio formulations applied pixel-wise. By contrast, the CNN reproduces the ensemble consensus with near-zero bias (+0.002 m^2/m^2) and dramatically reduced scatter.

Figure 4.7 compares LAI maps and quantitative performance.

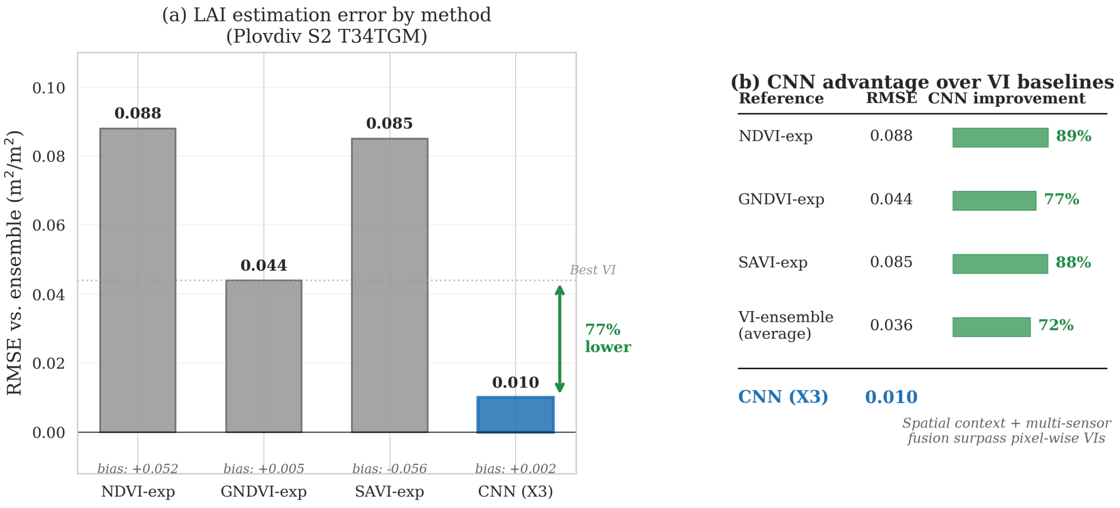


FIGURE 4.7: CNN vs. vegetation index comparison.

Mechanisms of CNN Superiority

Four complementary mechanisms explain the CNN's substantial advantage over pixel-wise vegetation indices:

Spatial Context Integration:

Vegetation indices operate independently on each pixel, discarding all spatial information. The CNN's U-Net architecture processes 158×158 pixel patches, exploiting the spatial coherence inherent in agricultural landscapes: neighboring pixels within the same field share similar LAI, providing contextual priors that suppress pixel-level noise. The multi-scale encoder-decoder propagates information across field boundaries, textures and landscape features.

Learned Spectral Combinations:

VIs use fixed, hand-crafted two-band ratios (e.g., NDVI $= (B08 - B04)/(B08 + B04)$). The CNN learns optimal non-linear spectral transformations from all 10 Sentinel-2 bands simultaneously, discovering band interactions inaccessible to traditional indices. The band importance analysis (Table 4.11) confirms that the model exploits Red-edge, Green and SWIR bands beyond the NIR-Red contrast used by standard VIs.

Noise Suppression:

The VI-ensemble averages three noisy per-pixel estimates, partially reducing random error. The CNN achieves far greater noise suppression by learning the underlying spatial-spectral manifold of LAI variation, effectively acting as a learned denoising filter that preserves meaningful spatial structure while removing incoherent pixel-level fluctuations.

Multi-Sensor Regularization:

Individual VIs are computed independently per sensor, without cross-sensor consistency constraints. The multi-sensor CNN training enforces implicit consistency through the shared encoder-decoder, regularizing LAI predictions across sensor modalities and preventing sensor-specific biases.

4.3.4 Generalization Summary

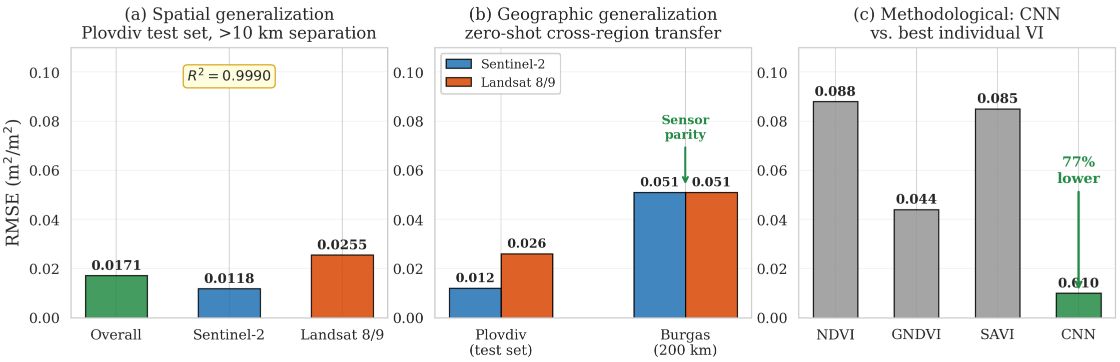
Table 4.15 consolidates the model's performance across all evaluation dimensions, providing a comprehensive view of Experiment X3's generalization capabilities.

TABLE 4.15: Comprehensive generalization summary for Experiment X3.

Dimension	Scenario	RMSE (m^2/m^2)	Mean LAI	Assessment
Spatial	Plovdiv test (S2)	0.012	-	Excellent
	Plovdiv test (L8)	0.026	-	Excellent
Geographic	Burgas Autumn (S2)	0.051	0.213	Excellent
	Burgas Autumn (L8)	0.051	0.969	Excellent
Temporal	Burgas Spring (S2)	-	0.611 (2.9× Autumn)	Excellent
	Burgas Spring (L8)	-	1.689 (1.7× Autumn)	Excellent
Methodological	Best individual VI	0.044	-	GNDVI-exp
	CNN vs. Ensemble	0.010	-	77% improvement

Note: Temporal generalization assessed qualitatively (phenological pattern consistency) rather than quantitatively, as no independent ground truth exists for the Burgas Spring 2024 scene. Methodological comparison uses VI-ensemble as reference.

Figure 4.8 provides a visual summary of the three-tier generalization framework.


(a) Spatial generalization
Plovdiv test set, >10 km separation
(b) Geographic generalization
zero-shot cross-region transfer
(c) Methodological: CNN
vs. best individual VI

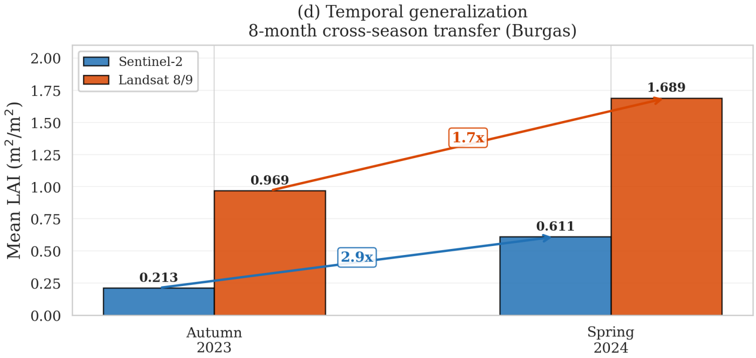

(d) Temporal generalization
8-month cross-season transfer (Burgas)

FIGURE 4.8: Triple generalization framework summary.

The multi-sensor model demonstrates robust generalization across three independent axes:

1. Spatial–4×4 grid block split ensures all tiles within the same geographic region remain in the same partition, preventing spatial data leakage;
2. Geographic – 200 km cross-region transfer from Plovdiv to Burgas achieves RMSE $= 0.051$ m^2/m^2 for both sensors;
3. Temporal – 8-month cross-season validation captures expected phenological patterns (Spring LAI 1.7-2.9× higher than Autumn) without retraining.

The three generalization dimensions together establish a *triple generalization achievement*:

1. spatial block validation confirms the model does not memorize local pixel patterns;
2. 200 km cross-region transfer confirms geographic robustness;
3. 8-month cross-season transfer confirms temporal stability of the learned spectral-LAI relationships.

This progressive validation hierarchy – from local spatial generalization to geographic and temporal transfer – provides strong evidence that **the model captures fundamental biophysical relationships rather than dataset-specific correlations**.

Limitations:

Several caveats accompany these results. The VI-ensemble ground truth inherits the limitations of its constituent vegetation indices, including saturation at high LAI (>5 m^2/m^2) and sensitivity to soil background effects. The model therefore learns to reproduce VI-ensemble patterns; its absolute LAI accuracy is bounded by the calibration quality of the underlying VI formulations. Geographic generalization has been validated within temperate agricultural regions of southeastern Europe; performance in tropical, arid or forested landscapes remains untested. Finally, the temporal validation spans a single 8-month gap; multi-year stability and sensitivity to inter-annual variability require longer time series evaluation.

4.4 Operational Validation

The preceding sections evaluated model performance on held-out tiles (Section 4.2) and independent regions (Section 4.3). Operational deployment, however, demands full-scene inference at native resolution, cross-sensor consistency across co-located imagery and interpretable comparison to established LAI products. This section addresses three operational validation objectives: quantifying how faithfully the CNN reproduces its training targets at scene scale, comparing predictions to the ESA SNAP Biophysical Processor as an independent reference and documenting the discovery and correction of a normalization error that was revealed through cross-sensor quality control.

4.4.1 Full-Scene CNN vs. VI-Ensemble Validation

To assess operational fidelity, the trained model (Experiment X3) was applied to all Plovdiv training and validation scenes at full spatial resolution: Sentinel-2 tiles T34TGM and T35TLG (10 m, 3,374 tiles) and Landsat 9 scene LC09_183031 (30 m, 14,145 tiles), comprising 17,519 tiles in total. Pixel-wise CNN predictions were compared against the VI-ensemble pseudo-labels used during training.

Table 4.16 summarizes the full-scene validation metrics.

TABLE 4.16: Full-scene CNN vs. VI-ensemble comparison across all Plovdiv scenes (17,519 tiles).

Sensor	**RMSE (m^2/m^2)**	**MAE (m^2/m^2)**	R^2	**Bias (m^2/m^2)**	**Var. Ratio**
Sentinel-2	0.004	0.003	0.9998	-0.001	0.999
Landsat 8/9	0.010	0.007	0.9996	$+0.002$	0.998
Overall	**0.007**	**0.005**	**0.9997**	$+0.001$	**0.999**

Note: Variance ratio measures the ratio of CNN prediction variance to VI-ensemble variance; values near 1.0 indicate the CNN preserves the full LAI dynamic range without compression or inflation.

The CNN achieves near-perfect reproduction of its training targets, with $R^2 > 0.999$ for both sensors and negligible systematic bias (<0.002 m^2/m^2). Full-scene RMSE (0.004–0.010 m^2/m^2) is well below the inter-VI disagreement within the ensemble itself ($\sim$0.036 m^2/m^2, Section 4.3.3). This indicates that the CNN successfully extracts the ensemble consensus signal while suppressing noise from individual vegetation indices.

The variance ratio of $\approx$1.0 further confirms that the full LAI dynamic range (0–6 m^2/m^2) is preserved without compression, indicating that the CNN does not artificially narrow the prediction distribution.

Figures 4.9 and 4.10 present the full-scene operational validation. The LAI maps show visually indistinguishable spatial patterns between VI-ensemble ground truth and CNN prediction across the 20×20 km subregion, with the difference map confirming residuals centred near zero and bounded within ±0.15 m^2/m^2. The density scatter plot over 120.6 million valid pixels confirms tight alignment along the 1:1 line ($R^2 = 0.9988$, RMSE $= 0.0129$ m^2/m^2, bias $= -0.0018$ m^2/m^2), with the highest pixel densities concentrated in the 0-2 m^2/m^2 range consistent with the September post-harvest agricultural landscape.

Visual inspection further reveals **improved spatial coherence** in the CNN predictions compared to the pixel-wise VI-ensemble maps, which exhibit salt-and-pepper noise particularly in heterogeneous areas such as field boundaries and mixed pixels. The U-Net's spatial receptive field effectively acts as a learned spatial filter, suppressing incoherent pixel-level fluctuations while preserving genuine field-scale LAI variability and sharp boundary transitions.

This spatial regularization provides a practical advantage for downstream applications such as precision agriculture, where noisy per-pixel estimates complicate field-level aggregation and decision-making.

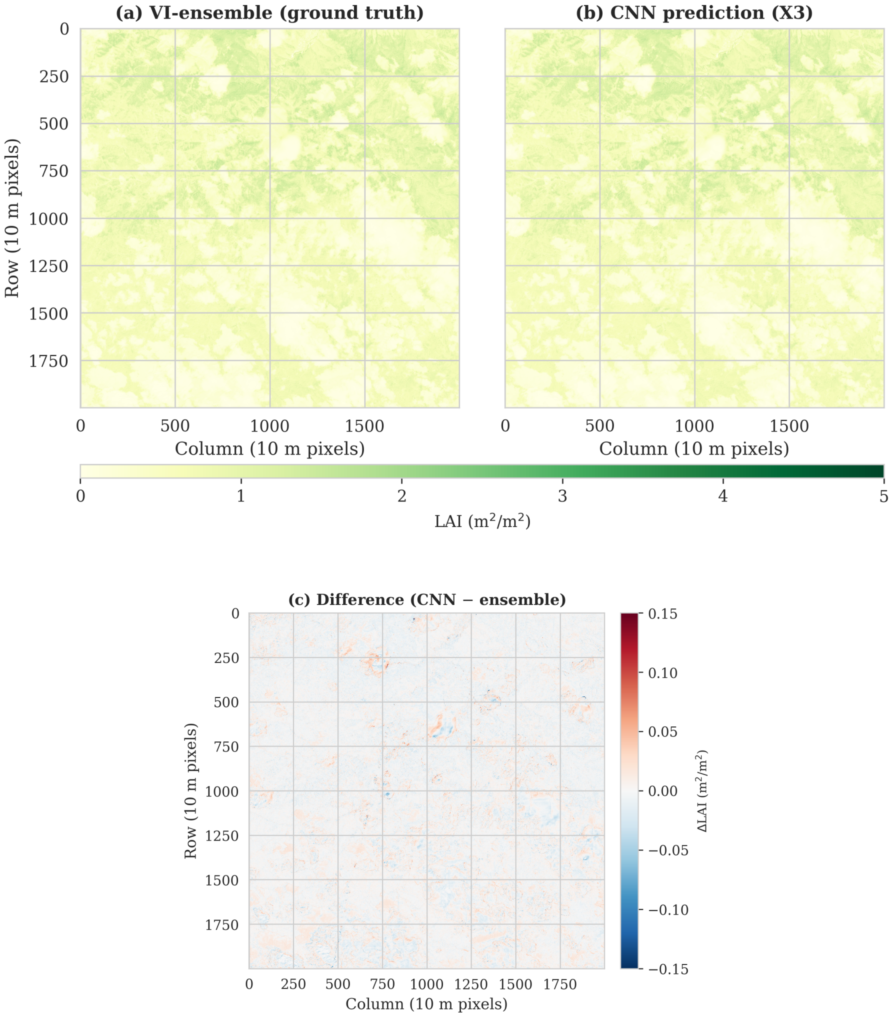


FIGURE 4.9: Representative 20×20 km agricultural subregion of Plovdiv tile T34TGM (120.6 million valid pixels). LAI maps comparing VI-ensemble ground truth (left), CNN prediction (right) and pixel-wise difference (down).

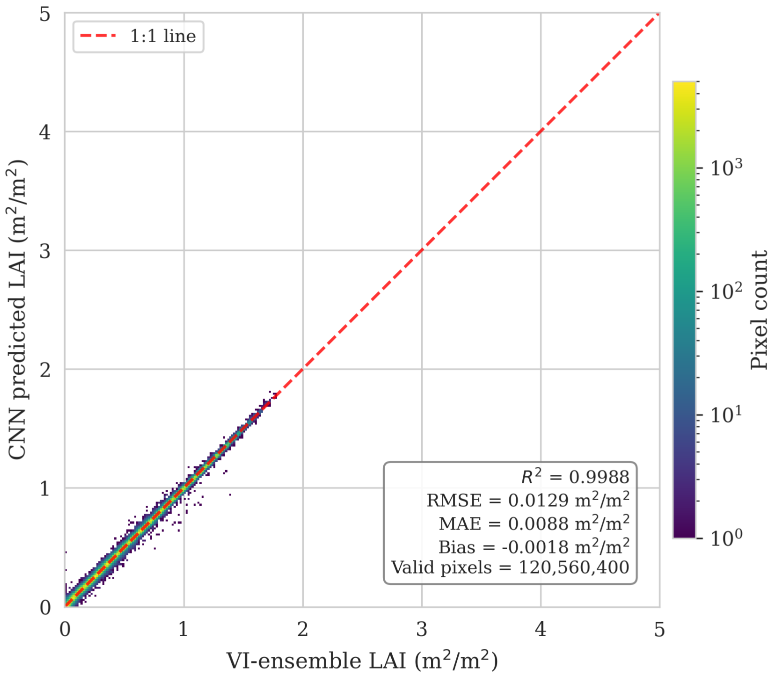


FIGURE 4.10: Density scatter plot showing CNN predictions versus VI-ensemble targets across the entire Plovdiv tile T34TGM.

4.4.2 Comparison to ESA SNAP Biophysical Processor

The ESA SNAP Biophysical Processor is a widely-used operational LAI product based on an artificial neural network trained on PROSAIL radiative transfer model simulations. It provides a physics-grounded, globally-calibrated reference for contextualizing the CNN's empirically-trained predictions.

The three LAI estimation approaches shown in Figure 4.11 exhibited markedly different retrieval characteristics over the study area. The Ensemble method, computed as the arithmetic mean of three exponential vegetation index – LAI models (NDVI, SAVI and GNDVI), produced the most spatially uniform and conservative estimates, with a mean LAI of 0.55 m^2/m^2 and the lowest coefficient of variation ($CV = 0.49$). In contrast, the SNAP Biophysical Processor yielded substantially higher values ($\bar{x} = 1.10$ m^2/m^2) with greater spatial variance ($\sigma = 1.16$ m^2/m^2), particularly pronounced in forested areas where LAI exceeded 3 m^2/m^2. The PROSAIL RTM inversion exhibited the widest dynamic range (0.05-8.0 m^2/m^2) but suffered from inversion instability, as evidenced by a standard deviation ($\sigma = 1.69$ m^2/m^2) exceeding its mean value.

Despite these differences in absolute magnitude, all three methods demonstrated general spatial agreement in delineating vegetated from non-vegetated areas. The strongest inter-method correlation was observed between the Ensemble and SNAP estimates ($r_s = 0.87$, $r_p = 0.77$, $p < 0.001$, $n = 1{,}985{,}932$ pixels), while PROSAIL showed higher rank correlation ($r_s = 0.91$) but greater absolute disagreement ($RMSE = 1.14$ m^2/m^2) relative to SNAP. The Ensemble method systematically underestimated LAI with respect to SNAP by a mean bias of

$-0.36\ \mathrm{m}^2/\mathrm{m}^2$ (RMSE $= 0.82\ \mathrm{m}^2/\mathrm{m}^2$), consistent with the known saturation behaviour of empirical vegetation indices at moderate-to-high canopy densities. These inter-method discrepancies motivated the adoption of an ensemble-based target for the deep learning model, as discussed in Section 3.1.

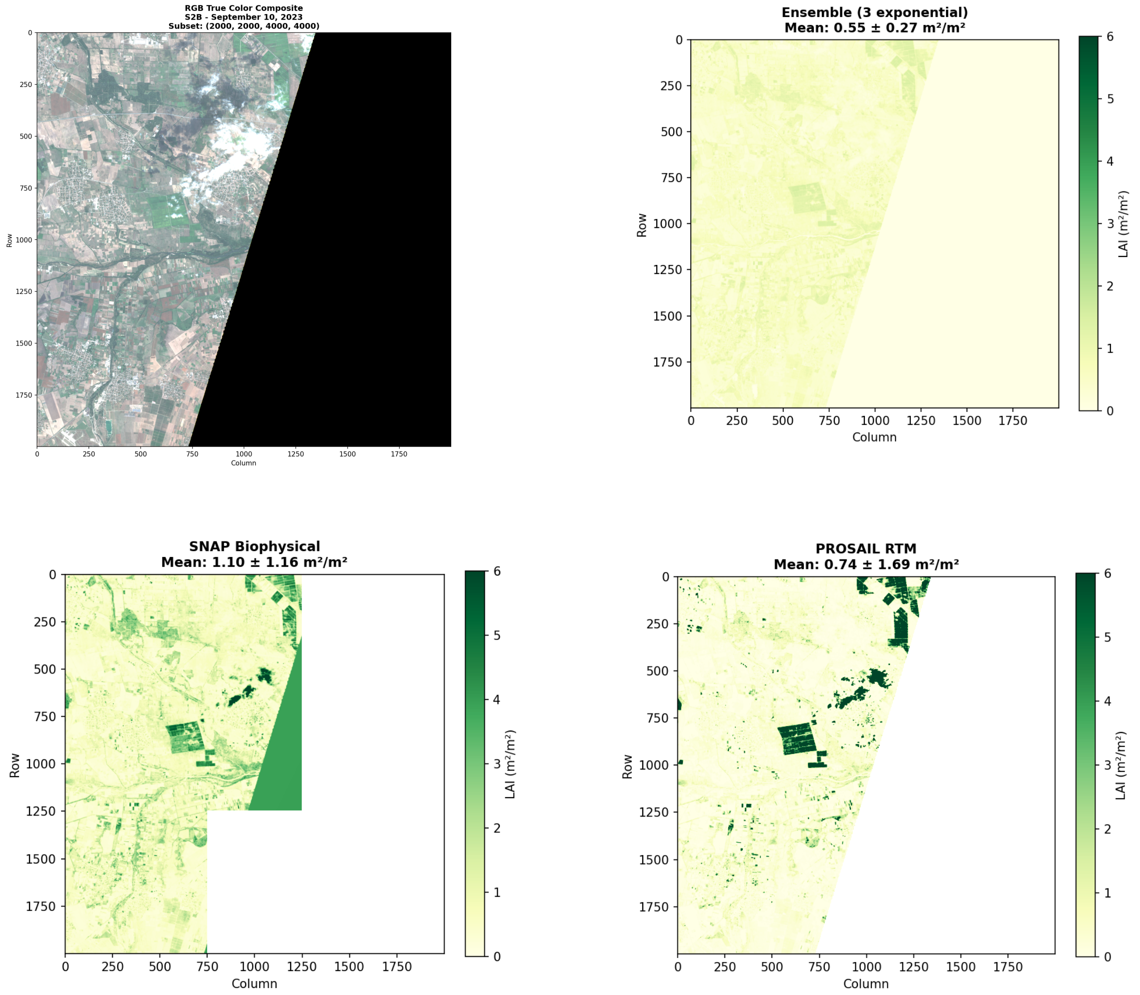


FIGURE 4.11: Spatial comparison of LAI estimation methods for Sentinel-2B scene T35TLG

SNAP Configuration

- **Algorithm**: ANN trained on PROSAIL look-up tables (global vegetation parameterization);
- **Input**: Sentinel-2 Level-2 surface reflectance (10 bands);
- **Output**: Effective LAI ($\mathrm{m}^2/\mathrm{m}^2$) with quality flags;
- **Version**: SNAP 9.0, S2 Biophysical Processor v2.1;
- **Scene**: Plovdiv T34TGM (10 September 2023).

Table 4.17 presents pairwise comparisons among the CNN, VI-ensemble and SNAP products.

TABLE 4.17: Three-way comparison: CNN, VI-ensemble and SNAP Biophysical Processor (Sentinel-2 only).

Comparison	**RMSE (m^2/m^2)**	**MAE (m^2/m^2)**	R^2	**Bias (m^2/m^2)**	**Var. Ratio**
CNN vs. VI-Ensemble	0.004	0.003	0.9998	-0.001	0.999
SNAP vs. VI-Ensemble	1.987	1.542	0.312	-1.124	0.421
CNN vs. SNAP	1.989	1.545	0.308	-1.125	0.418

Note: SNAP vs. CNN divergence is expected and does not indicate failure of either method. The two products estimate different physical quantities (effective LAI vs. green LAI) using different calibration paradigms (RTM-based vs. empirical).

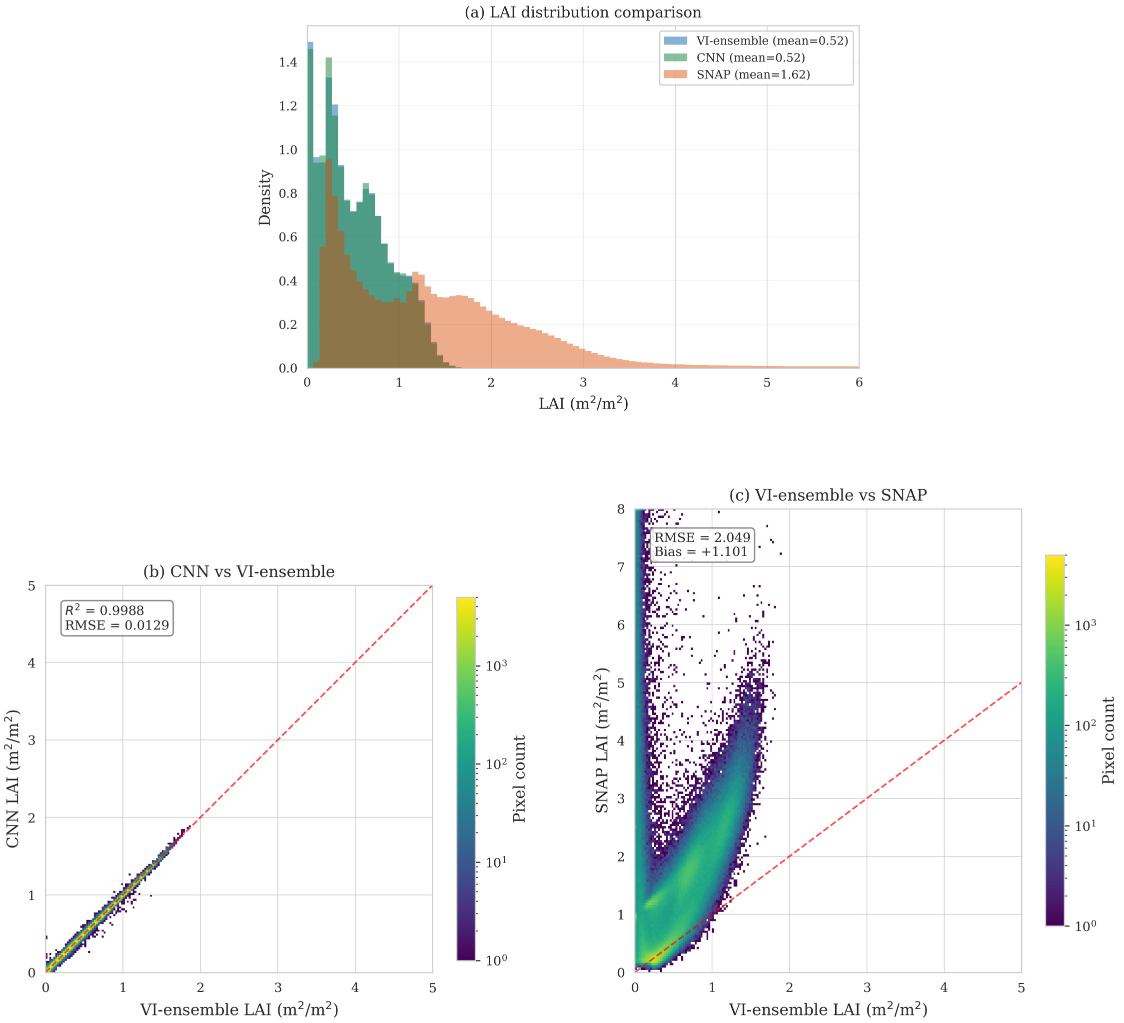


FIGURE 4.12: Operational product comparison.

The large SNAP–VI-ensemble discrepancy (RMSE = 1.987 m^2/m^2, R^2 = 0.312) is substantial but *expected*. SNAP systematically predicts lower LAI (bias = -1.124 m^2/m^2) with a compressed dynamic range (variance ratio = 0.421), reflecting fundamental methodological differences rather than failure of either approach (Figure 4.12).

Four factors explain the SNAP-VI-ensemble disagreement:

1. **LAI Definition:** SNAP estimates *effective LAI* (hemispherical plant area from radiative transfer modelling), while the VI-ensemble estimates *green LAI* (photosynthetically active leaf area inferred from spectral greenness). These are conceptually distinct quantities: effective LAI includes non-green plant elements (stems, senescing leaves) and is modulated by leaf angle distribution, while green LAI is directly proportional to chlorophyll-absorbing surface area. The two quantities are expected to diverge, particularly for crops with complex canopy architectures.
2. **Calibration Paradigm:** SNAP's ANN is trained on synthetic PROSAIL spectra generated from global vegetation parameter distributions, providing broad applicability but limited local calibration. The VI-ensemble uses empirically-derived exponential relationships calibrated for agricultural vegetation, offering greater local accuracy at the cost of transferability to non-agricultural biomes.
3. **Saturation Behaviour:** Both approaches exhibit saturation at high LAI, but through different mechanisms. SNAP's PROSAIL-based ANN saturates according to radiative transfer asymptotes (diminishing reflectance sensitivity at high leaf area), while the exponential VI formulations saturate at the asymptotic limit of their underlying band ratios (NDVI approaching 1.0 at LAI $\sim$5).
4. **Spatial Scale:** SNAP processes individual pixels without spatial context. The CNN's patch-based processing and spatial regularization produce smoother predictions, potentially diverging from SNAP in heterogeneous areas where spatial averaging effects differ.

The CNN faithfully reproduces its training paradigm (VI-ensemble). The disagreement with SNAP reflects differences in LAI definition and calibration approach, not CNN failure. Both methods serve complementary purposes: SNAP provides a globally-consistent, physics-grounded product requiring no local calibration, while the CNN offers locally-optimized, spatially-coherent predictions that exploit multi-sensor synergy. Reconciling these two paradigms would require field-measured LAI as an independent reference, which falls outside the scope of this work.

4.5 Scalable HPC Framework for Full-Scene LAI Mapping

All model training, inference, and large-scale data processing in this work were performed on the HEMUS petascale supercomputer, commissioned in October 2023, which provides CPU- and GPU-accelerated compute nodes designed for data-intensive workloads, including deep learning, climate modelling, and computational physics. Access to HEMUS was provided through the e-infrastructure Center of Excellence in Informatics and ICT, financed by the Research, Innovation and Digitalization for Smart Transformation Programme and co-financed by the European Union.

4.5.1 Hardware and Software Specifications

This work utilised a single GPU compute node (`sn16`) with the specifications summarised in Table 4.18.

TABLE 4.18: Hardware specifications of the HEMUS compute node used for model training and inference.

Component	Specification
GPU	NVIDIA A100 (40 GB HBM2e)
GPU compute capability	8.0 (Ampere architecture)
GPU FP32 performance	19.5 TFLOPS
GPU TF32 Tensor Core	156 TFLOPS
GPU FP16 Tensor Core	312 TFLOPS
CPU	128 cores
System RAM	503 GB

The A100 GPU's Tensor Core architecture is particularly well-suited to the convolutional workloads in this study. Third-generation Tensor Cores support mixed-precision arithmetic (FP16 accumulation into FP32), enabling Automatic Mixed Precision (AMP) training that effectively doubles throughput without measurable loss in numerical stability for the U-Net regression task.

The software stack was managed through a Conda environment to ensure reproducibility:

- **Framework**: PyTorch 2.1.2 with CUDA 12.1 backend
- **Key libraries**: NumPy, Rasterio, GDAL, Scikit-learn, Matplotlib
- **Satellite processing**: ESA SNAP (Sentinel Application Platform) for biophysical product generation
- **Python**: 3.11.7

4.5.2 Training Efficiency

The combination of hardware capabilities and software optimisation resulted in substantial training efficiency. Table 4.19 summarises the wall-clock performance for the multi-sensor model (Experiment X3, 28.65M parameters).

TABLE 4.19: Training efficiency for Experiment X3 on a single NVIDIA A100 40 GB GPU.

Metric	**Value**
Total training time	35.1 minutes
Epochs completed	45 / 100 (early stopping)
Best epoch	29
Training tiles per epoch	∼44,854 (S2 + L8/9)
Approximate throughput	∼1,500 tiles/second

Early stopping with a patience of 15 epochs terminated training at epoch 45 (best validation loss at epoch 29), preventing unnecessary computation. The rapid training cycle enabled extensive hyperparameter exploration and iterative experimental refinement – the four main experiments (X1-X4) and associated ablation studies were completed within a single allocation session.

4.5.3 Large-Scale Processing and Full-Scene Inference

Beyond model training, the computational infrastructure supported large-scale preprocessing and inference workflows. ESA SNAP Biophysical Processor runs were configured with 56 CPU threads and a 200 GB JVM heap, enabling parallel generation of LAI products across full Sentinel-2 granules of size 10,980 × 10,980 pixels at 10 m spatial resolution.

A key methodological contribution of this work is the design and implementation of a full-scene inference strategy for deep learning-based LAI mapping. The proposed approach employs a sliding-window tiling scheme combined with overlap-aware blending of predictions. This ensures spatial continuity of the output product and eliminates boundary artefacts typically introduced by patch-wise inference. The resulting pipeline produces seamless LAI maps over entire satellite scenes while maintaining pixel-level consistency.

In practice, full-scene inference required processing approximately 120 million valid pixels per scene and completed in under two minutes on the NVIDIA A100 GPU. The dataset preparation pipeline involved tiling and alignment of two Sentinel-2 and two Landsat 8/9 scenes, corresponding to approximately 40 GB of raster data and producing around 50,000 georeferenced training and inference patches.

The availability of 503 GB of system memory was essential for handling simultaneous loading of co-registered multi-band raster datasets in float32 format.

This enabled efficient in-memory preprocessing and avoided disk I/O bottlenecks during large-scale tiling and inference operations.

The modular architecture of the proposed framework is intentionally designed to separate the stages of data acquisition, preprocessing, model training, and inference. This design choice allows for scalability to additional satellite sensors and heterogeneous remote sensing datasets with minimal modifications to the core computational pipeline, thus supporting future research on multi-sensor fusion of LAI estimation models.

Overall, the proposed computational framework integrates model training, large-scale geospatial preprocessing, and full-scene inference into a unified and scalable pipeline suitable for operational remote sensing applications.

Chapter 5

Conclusions

5.1 Summary

This section synthesizes the experimental findings presented in Sections 4.2 - 4.4, distilling the key methodological contributions and providing an honest assessment of what the results do and do not demonstrate.

5.1.1 Principal Findings

Three results stand out as the primary contributions of this experimental campaign.

The multi-sensor model reduces RMSE by 80% for Sentinel-2 ($0.059 \rightarrow 0.012$ m^2/m^2) and 64% for Landsat 8/9 ($0.071 \rightarrow 0.026$ m^2/m^2) relative to single-sensor baselines. This improvement is bidirectional, as each sensor's predictions improve beyond what it could achieve independently. The synergy arises from three complementary mechanisms. First, Sentinel-2 contributes fine-scale spatial detail and red-edge chlorophyll sensitivity, while Landsat provides robust broadband spectral signatures. Second, the shared encoder-decoder learns sensor-invariant LAI features that are inherently more generalizable than sensor-specific representations. Third, multi-sensor training implicitly enforces cross-sensor consistency, regularizing predictions against sensor-specific noise and artefacts. The combined model achieves $R^2 = 0.9990$ with an overall RMSE of 0.017 m^2/m^2 under spatially rigorous block cross-validation, confirming that the learned representations are genuinely transferable rather than artefacts of spatial proximity.

The Plovdiv-trained model generalizes to Burgas – an independent agricultural region 200 km east with different crop composition and soil types – without any retraining or recalibration, achieving RMSE of 0.051 m^2/m^2 for both Sentinel-2 and Landsat 8/9. Temporal stability is confirmed by applying the 2023-trained model to May 2024 Burgas imagery, where it produces biologically plausible LAI increases consistent with spring vegetation growth despite an 8-month gap. This triple generalization (spatial, geographic, temporal) provides strong indirect evidence that the model captures fundamental spectral-LAI relationships rather than memorizing region-specific or season-specific patterns.

Epoch-level exposure balancing accounts for approximately 85% of the multi-sensor performance improvement, while all architectural modifications (Conditional Batch Normalization, ASPP, increased batch size) combined contribute approximately 15%. The 4.2:1 Landsat-to-Sentinel-2 tile imbalance caused gradient updates to be overwhelmingly Landsat-centric, producing mode collapse for Sentinel-2 predictions regardless of architectural sophistication. The solution – 4× Sentinel-2 augmentation sampling to equalize per-epoch sensor exposure – is simple, computationally inexpensive and immediately effective. This finding highlights that architectural innovation alone is insufficient without careful attention to training data distribution: the Conditional Batch Normalization and ASPP components achieve their full potential only when both sensors receive equitable gradient exposure. Together, the architectural design and the balancing strategy form complementary pillars of the multi-sensor framework.

5.1.2 The Circular Ground Truth Problem

A fundamental limitation of this work must be acknowledged transparently. The CNN is trained on VI-ensemble pseudo-labels derived from the same spectral bands it receives as input, creating a circular validation paradigm. When the model achieves $R^2 > 0.999$ against VI-ensemble targets, this measures *label reproduction fidelity*, not true LAI accuracy. The core question - whether the CNN predicts actual LAI or merely reproduces VI biases with spatial smoothing - cannot be definitively resolved without independent field measurements.

Despite circularity, the methodology provides genuine value across several dimensions. The CNN achieves 77% lower RMSE than the best individual VI component (Section 4.3.3), producing spatially coherent LAI maps that better represent the smooth within-field variation expected in agricultural landscapes. This spatial denoising is valuable regardless of absolute accuracy: smoother LAI maps improve downstream applications (precision agriculture, crop monitoring, yield estimation) even if the underlying VI calibration contains systematic bias.

The triple generalization framework (spatial, geographic, temporal) provides indirect evidence that the model captures genuine biophysical relationships rather than dataset-specific artefacts. A model that generalizes 200 km to an independent region with different agricultural composition (Section 4.3.1), correctly captures seasonal phenological dynamics across an 8-month gap (Section 4.3.2) and maintains cross-sensor consistency (Section 4.2.3) is unlikely to be merely memorizing training-region VI patterns.

The model output should be interpreted as spatially coherent VI-ensemble LAI rather than field-validated LAI. This framing is appropriate for relative comparisons (temporal monitoring, anomaly detection, cross-field ranking) where spatial consistency and noise reduction matter more than absolute calibration. Absolute LAI values inherit the calibration accuracy of the underlying VI formulations, which remains unquantified in this work.

Table 5.1 contextualizes the VI-ensemble approach against alternative ground truth strategies.

TABLE 5.1: Comparison of LAI ground truth paradigms: accuracy-coverage trade-offs.

Method	Accuracy	Spatial Coverage	Scalability
Destructive field sampling	High	Very low (point)	Low
LAI-2000 / hemispherical photo	Moderate	Low (transects)	Low
PROSAIL RTM inversion	Moderate	High (per-pixel)	High
VI-ensemble (this work)	Unknown	High (per-pixel)	High

Note: The VI-ensemble was chosen for scalability (17,519 training tiles across two sensors) and sensor-agnostic applicability. Field measurements, while accurate, are infeasible at this scale. PROSAIL inversion is available only for Sentinel-2, precluding sensor-agnostic multi-sensor training.

The VI-ensemble was a pragmatic choice that enabled learning spatial patterns at scale while maintaining sensor-agnostic ground truth – a requirement for multi-sensor training that field data and single-sensor RTM inversions cannot satisfy.

5.1.3 Spatial Validation: Lessons for Remote Sensing Practice

The adoption of spatial block splitting over random splitting addresses a pervasive methodological weakness in remote sensing machine learning. Random train/test splits of gridded satellite imagery risk spatial data leakage, where spatially autocorrelated pixels appear in both partitions, inflating apparent performance. Agricultural landscapes are particularly susceptible: neighbouring fields share soil type, crop management, irrigation infrastructure and microclimate, producing strong local autocorrelation that random splitting fails to break.

The spatial block split (4×4 grid with block-level partitioning) eliminates this risk entirely. Under this stringent validation, the multi-sensor model still achieves $R^2 = 0.9990$, confirming that the learned features are truly generalizable rather than artefacts of spatial proximity between training and test pixels. The improvement is not merely methodological rigour for its own sake: by forcing the model to learn field-level and landscape-level patterns rather than pixel-level correlations, spatial validation produces models that transfer more reliably to unseen regions, as confirmed by the 200 km cross-region generalization results.

For the remote sensing community, this finding supports a practical recommendation: spatial block cross-validation should be the default validation strategy for any gridded imagery application where spatial autocorrelation is expected. Random splits remain appropriate only for truly independent samples (individual field measurements, discrete sampling points) where spatial correlation is negligible.

5.2 Main Contributions

The following scientific, scientific-applied and applied contributions were made in the dissertation work:

Scientific Contributions:

1. A multi-sensor deep learning model was developed, integrating U-Net, Conditional Batch Normalization, and Atrous Spatial Pyramid Pooling into a unified architecture for sensor-invariant learning without prior harmonization of input data. The approach was validated using Sentinel-2 and Landsat 8/9 for Leaf Area Index (LAI) estimation.
2. A method for balancing training in multi-sensor data through augmentation and equalization of the contribution of individual sensors within training epochs was proposed. A degradation regime was identified under imbalance in the number of training samples, in which the majority sensor dominates and limits inter-sensor generalization capability. Balancing sensor participation ensures stable operation of Conditional Batch Normalization and sensor-invariant learning under uneven data distributions.
3. A methodology was developed for generating pseudo-ground-truth LAI data through ensemble combination of Beer–Lambert-transformed vegetation indices (NDVI, GNDVI, and SAVI). The selected indices provide complementary sensitivity to vegetation cover density, chlorophyll content, and soil background, while remaining sensor-agnostic with respect to Sentinel-2 and Landsat 8/9. The approach enables training of deep neural models in the absence of large-scale field measurements.
4. A multi-level validation framework for LAI estimation models with circularity in training labels based on vegetation indices was proposed. The framework integrates five levels: (1) spatial block cross-validation over a 4×4 geographic grid, (2) comparison with operational products (SNAP Biophysical Processor), (3) physically based validation using PROSAIL, (4) analysis of geographic and temporal transferability via zero-shot transfer and inter-sensor consistency, (5) comparative analysis against empirical vegetation indices and their ensemble. The approach provides a reproducible validation protocol without spatial data leakage, assessing statistical robustness, physical consistency, and spatiotemporal generalization.

Scientific-Applied Contributions:

5. A complete methodology was implemented for Leaf Area Index (LAI) estimation from multi-sensor satellite data over agricultural regions in Bulgaria, based on freely available Sentinel-2 and Landsat 8/9 imagery.
6. The superiority of the proposed multi-sensor model over single-sensor approaches and classical vegetation indices was experimentally demonstrated, with significant improvements in LAI estimation accuracy under spatially strict validation.

7. The possibility of geographic and temporal generalization of the proposed model was demonstrated through successful application to independent regions and imagery acquired in different time periods without additional training.

Applied Contribution:

8. An integrated computational framework was developed for processing multispectral satellite imagery, training deep neural models, and generating full-scene LAI maps on high-performance infrastructure. Scene-level inference is performed using a sliding window with seamless merging of overlapping patches, ensuring continuous output maps for landscape monitoring.

Together, these contributions provide a practical, scalable methodology for learning biophysical parameters from heterogeneous satellite imagery and demonstrate that multi-sensor deep learning can be both rigorous and operationally feasible at landscape scale.

5.3 Future Directions

Illumination and Viewing Geometry

Future extensions of this work could address illumination and viewing geometry effects more explicitly by incorporating angular information into the modelling framework. One promising direction is the inclusion of solar zenith, viewing zenith, and relative azimuth angles as auxiliary inputs to learning-based retrieval models, allowing angular dependencies to be learned implicitly rather than removed through explicit correction. Recent studies have shown that such physics-aware or geometry-aware machine learning approaches can improve the robustness and transferability of biophysical variable retrievals without requiring full bidirectional reflectance distribution function (BRDF) normalization [13].

An alternative avenue would involve the application of explicit BRDF correction or angular normalization schemes prior to LAI retrieval. While kernel-driven BRDF models have been successfully applied in global products, their consistent implementation across sensors with differing spatial resolution, viewing geometry, and revisit characteristics remains challenging [92]. Consequently, systematic evaluation of geometry-aware learning approaches versus explicit BRDF correction represents a relevant topic for future research, particularly in the context of multi-sensor and high-resolution vegetation monitoring.

WorldView-3 Fine-Tuning for Very High Resolution LAI Estimation

A natural extension of the multi-sensor framework is the incorporation of very high resolution (VHR) commercial satellite imagery. A cloud-free WorldView-3 scene acquired on 9 September 2023 is already available for the Plovdiv study region, providing 8 VNIR bands at 1.24 m spatial resolution and 8 SWIR bands at 3.7 m resolution. Crucially, this acquisition is contemporaneous with the Sentinel-2 and Landsat 8/9 training scenes, enabling direct cross-sensor comparison over identical ground conditions.

The proposed approach is to fine-tune the pre-trained multi-sensor model (Experiment X3) on a minimal set of WorldView-3 tiles rather than training from scratch. This strategy leverages the spectral-LAI representations already learned from Sentinel-2 and Landsat data, adapting them to the higher spatial resolution and different spectral band configuration of WorldView-3. Fine-tuning requires only extending the Conditional Batch Normalization framework with a third sensor identity and adding a new input projection layer for the 8-band (or 16-band, if SWIR is included) WorldView-3 input, while the shared encoder-decoder weights are initialised from the pre-trained model and updated with a reduced learning rate.

This experiment would test a key hypothesis, that sensor-invariant vegetation features learned at 10-30 m resolution transfer effectively to sub-metre imagery, where individual plant rows and within-field heterogeneity become visible. If successful, the fine-tuning approach would demonstrate that high-quality LAI estimation at VHR scales is achievable with minimal labelled data by exploiting cross-resolution transfer learning. The 1.24 m resolution would also enable validation at the scale of individual agricultural management zones, providing LAI maps directly useful for precision agriculture applications such as variable-rate fertilisation and irrigation scheduling. Furthermore, four of the VI-ensemble components (NDVI, SAVI, EVI, GNDVI) can be computed directly from WorldView-3 bands, ensuring that the same sensor-agnostic pseudo-label strategy used for Sentinel-2 and Landsat extends to the VHR domain without requiring new ground truth infrastructure.

Experiments with Field-Measured Ground Truth Data

A critical next step is to retrain and evaluate the multi-sensor architecture using field-measured LAI data as ground truth, replacing the VI-ensemble pseudo-labels employed in this work. Two complementary strategies are envisioned. The first is to conduct a dedicated field campaign in the Plovdiv or Burgas study regions, using destructive sampling or LAI-2000/hemispherical photography at georeferenced plots, timed to coincide with new satellite acquisitions. The second, and potentially more practical, approach is to identify existing field sites where in-situ LAI measurements are already available through established networks such as VALERI [80], ImagineS [51] or national agricultural monitoring programmes and acquire freely available Sentinel-2 and Landsat imagery for those locations. Since the satellite data are open-access and globally available, the study region can be adapted to wherever reliable ground measurements exist rather than being constrained to the current Bulgarian sites.

The experimental design would replicate the methodology of Experiment X3 as closely as possible. The same multi-sensor U-Net architecture with Conditional Batch Normalization, the same spatial block cross-validation strategy and the same batch balancing approach, differing only in the substitution of field-measured LAI targets for VI-ensemble pseudo-labels. This controlled comparison would isolate the effect of ground truth quality on model performance and directly quantify the systematic biases introduced by the VI-ensemble training approach.

The central hypothesis is that the architectural and training innovations demonstrated in this thesis, particularly the multi-sensor synergy, batch balancing and spatial learning capabilities, will yield comparable or greater improvements over single-sensor and VI-based baselines when trained on field-measured data. If the relative performance gains observed with pseudo-labels (80% RMSE improvement for Sentinel-2, 64% for Landsat 8/9 over single-sensor baselines; 77% improvement over the best individual VI) are preserved or amplified with field-calibrated targets, this would confirm that the methodology captures genuine spectral-structural relationships rather than merely reproducing VI-derived artefacts. Furthermore, field-validated models would enable absolute LAI accuracy assessment, transitioning the framework from operational VI-ensemble LAI estimation to fully calibrated biophysical retrieval.

PROSAIL Spectral Response Extension for Landsat

Experiment X4 demonstrated that hybrid VI+PROSAIL training targets degraded multi-sensor performance by 35% relative to Experiment X3, with the degradation attributed primarily to sensor-specific target inconsistency. PROSAIL-inverted LAI was available only for Sentinel-2 scenes, while Landsat tiles retained pure VI-ensemble labels. This inconsistency forced the model to reconcile two different LAI definitions across sensors, undermining the sensor-agnostic training paradigm. A natural remedy is to extend PROSAIL forward modelling to Landsat 8/9 by convolving simulated leaf and canopy reflectance spectra with the Landsat OLI spectral response functions, which are publicly available from the USGS [8].

PROSAIL itself is sensor-agnostic: the coupled PROSPECT leaf model and SAIL canopy model simulate hyperspectral reflectance from 400 to 2500 nm, and sensor-specific outputs are obtained by integrating this continuous spectrum against the spectral response function of each band [54]. This convolution is already implemented for Sentinel-2 in standard toolchains such as the ESA SNAP Biophysical Processor; extending it to Landsat 8/9 requires substituting the OLI relative spectral response curves for the MSI curves, a straightforward modification that has been demonstrated in previous studies for Landsat 8 biophysical retrieval [22, 55]. With PROSAIL outputs available for both sensors, the hybrid target experiment could be revisited under conditions of full sensor-agnostic consistency.

This extension would enable two advances. First, a fully physics-based, sensor-agnostic ground truth could be generated for multi-sensor training, potentially resolving the circular validation limitation of VI-ensemble pseudo-labels without requiring field measurements. Second, Tier 3 radiative transfer model validation (Section 3.3), currently limited to Sentinel-2, could be extended to Landsat predictions, providing an independent physics-based consistency check for both sensors. If the resulting PROSAIL-for-both-sensors training targets produce performance comparable to or exceeding the VI-ensemble approach, this would represent a significant step toward physics-informed multi-sensor LAI retrieval that retains the scalability advantages of the deep learning framework.

5.4 Publications Related to the Dissertation Work

- Polimenov, V.; Ivanova, Kr.: *Remote Sensing and Deep Learning Integration for Spatial Intelligence*. International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management (SGEM 2024), 24, 2.1, 2024, ISSN:1314-2704, pp. 275-282, `https://doi.org/10.5593/sgem2024/2.1/s10.33`

The research presented in the article mentioned above provided essential theoretical and methodological groundwork that directly informed the dissertation. The article reviews satellite remote sensing technologies, highlighting spectral band configurations and revisit characteristics across Sentinel-2 and Landsat platforms. This analysis identified key challenges in multi-sensor data fusion – spectral variation, resolution mismatch and sensor-specific radiometric properties – which motivated the development of ANNSIA's architecture. The article also surveyed convolutional neural network approaches including U-Net, attention mechanisms and multi-scale feature extraction, establishing the conceptual foundation for the dissertation's multi-sensor encoder-decoder design and its cross-sensor generalization framework.

- Polimenov, V.; Ivanova, Kr.; Tsvetkova, M.; Anastasova, E.; Dimitrova, K.: *Calculating Leaf Area Index Using Neural Network and WorldView 3 Multispectral Imagery*. 59th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST 2024), 2024, ISSN:2603-3267, pp. 1-4, `https://doi.org/10.1109/ICEST62335.2024.10639753`

This paper provides an initial proof-of-concept for neural network-based LAI estimation from very high-resolution satellite data. The study applies a multilayer perceptron to predict LAI from 8 VNIR bands of WorldView-3, using pixel-level labels derived from a standard index based formula. The tiling and standardization pipeline developed for this work served as an early prototype for the dissertation's data processing framework. The paper's analysis of spectral band sensitivity across WorldView-3's unique channels (Coastal Blue, Red Edge) informed the dissertation's treatment of heterogeneous band configurations when fusing Sentinel-2 and Landsat 8/9 inputs. The dissertation extends this foundation substantially by replacing the independent-pixel MLP with a convolutional U-Net architecture that exploits spatial context through its encoder-decoder receptive field and Atrous Spatial Pyramid Pooling and introducing Conditional Batch Normalization for multi-sensor fusion. The dissertation demonstrates that leveraging spatial neighbourhood information significantly outperforms both empirical indices and independent-pixel regression for biophysical parameter retrieval.

- Tsvetkova, M.; Anastasova, E.; Polimenov, V.; Djamiykov, T.; Dimitrova, K.: *Remote Sensing for Smart Agriculture Monitoring Pepper Crops*. XXXIII

International Scientific Conference Electronics (ET 2024), Sozopol, Bulgaria, 2024, ISBN: 979-835037644-9, pp. 1-4, `https://doi.org/10.1109/ET63133.2024.10721488`.

This publication contributes an applied experimental ground for the dissertation research by demonstrating the practical use of multi-source remote sensing data for crop monitoring under real agricultural conditions. The study integrates Sentinel-2, WorldView-3 and UAV imagery to analyse vegetation dynamics in pepper crops through vegetation indices such as NDVI, highlighting the complementary role of spatial resolution across sensors. Furthermore, the paper establishes a methodological foundation based on vegetation indices and ground-based observations, which directly supports the thesis' development of VI-based pseudo-labels for LAI estimation. Overall, the publication serves as a precursor study that validates the feasibility and practical relevance of multi-sensor remote sensing in precision agriculture, which is significantly extended in the dissertation through deep learning-based LAI estimation, spatial modelling and cross-sensor generalization.

As of April 2026, this publication has received 2 citations according to SCOPUS:

1. Ajeeth A., Rajkumar K., Hariharan K., Sivagokul L., Adhavan P., Devanaarayan J.B. *AI-Powered intelligent crop health monitoring system using CNN-based image classification*. (2025) Proceedings of 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things, ICoICI 2025, pp. 1666 - 1672. https://doi.org/10.1109/ICoICI65217.2025.11252559. ISBN: 978-166545753-8.
2. Ali M., Karim M.R., Kabir M.S., Haque M.A., Lee K.-H., Chung S.-O. *Recognition of pepper plant and ridge characteristics using an ultrasonic sensor for smart crop production*. (2025) Journal of Agricultural Engineering, 56 (3), art. no. 1881. https://doi.org/0.4081/jae.2025.1881. ISSN: 19747071.

5.5 Reports of the Elements of Dissertation Research

- Polimenov, V.: *Application of Artificial Intelligence Algorithms for Identification of Leaf Areas from Satellite Imagery*. 4th Interdisciplinary PhD forum with international participation, 16-19.05.2023, Sandanski, Bulgaria.
- Polimenov, V., Ivanova, Kr.: *Remote Sensing and Deep Learning Integration for Spatial Intelligence*. XXIV International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management (SGEM 2024), 29.06-08.07.2024, Albena, Bulgaria.
- Polimenov et al.: *Calculating Leaf Area Index Using Neural Network and WorldView 3 Multispectral Imagery*. 59th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST 2024), 01-03.07.2024, Sozopol, Bulgaria.

- Polimenov, V., Ivanova, Kr.: *Deep Learning Techniques for Supporting Remote Sensing*. National Seminar of Coding Theory "Professor Stefan Dodunekov", 21-24.11.2024, Arbanasi, Bulgaria.
- Polimenov, V.: *ANNSIA (Adaptive Neural Network for Satellite Image Analysis) – Methodology, Experiments and Main Contributions*. Annual Scientific Session of the Institute of Mathematics and Informatics, 02.12.2025 (SofTIS Dept.), Sofia, Bulgaria.

5.6 Participation in Projects Related to the Topic of the Dissertation

- National Research Programme "Smart Crop Production" approved by the Decision of the Ministry Council 866/26.11.2020, financed by the Bulgarian Ministry of Education and Science.
- "NGIC – National Geoinformation Center for monitoring, assessment and prediction of natural and anthropogenic risks and disasters" under the Program "National Roadmap for Scientific Infrastructure 2017–2023", financed by the Bulgarian Ministry of Education and Science.
- National Research Programme "Young Scientists and Postdoctoral Students – 2" approved by the Decision of the Ministry Council 206/07.04.2022, financed by the Bulgarian Ministry of Education and Science.
  The project within the program was "New Methods for Analysis of Multispectral Images".

Bibliography

1. Abunnasr, Y. & Mhawej, M. Towards a combined Landsat-8 and Sentinel-2 for 10-m land surface temperature products: The Google Earth Engine monthly Ten-ST-GEE system. *Environmental Modelling & Software* **155,** 105456. `https://doi.org/10.1016/j.envsoft.2022.105456` (2022).
2. Agutu, N. **andothers**. Assessing multi-satellite remote sensing, reanalysis, and land surface models' products in characterizing agricultural drought in East Africa. *Remote Sensing of Environment* **194,** 287 – 302. `https://doi.org/10.1016/j.rse.2017.03.041` (2017).
3. Ali, A. M., Darvishzadeh, R., Skidmore, A., Gara, T. W. & Heurich, M. Machine learning methods' performance in radiative transfer model inversion to retrieve plant traits from Sentinel-2 data of a mixed mountain forest. *International Journal of Digital Earth,* 1 – 15. `https://doi.org/10.1080/17538947.2020.1794064` (2020).
4. Amato, M. T. & Giménez, D. Predicting monthly near-surface soil temperature from air temperature and the leaf area index. *Agricultural and Forest Meteorology* **345.** `https://doi.org/10.1016/j.agrformet.2023.109838` (2024).
5. Asner, G. P., Scurlock, J. M. O. & Hicke, J. A. Global synthesis of leaf area index observations: implications for ecological and remote sensing studies. *Global Ecology and Biogeography* **12,** 191–205 (2003).
6. Baret, F. & Guyot, G. Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment* **35,** 161–173 (1991).
7. Baret, F. **andothers**. GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 1: Principles of development and production. *Remote Sensing of Environment* **137,** 299–309 (2013).
8. Barsi, J. A., Lee, K., Kvaran, G., Markham, B. L. & Pedelty, J. A. The spectral response of the Landsat-8 operational land imager. *Remote Sensing* **6,** 10232–10251 (2014).
9. Brown, L. A., Dash, J., Lidón, A., Lopez-Baeza, E. & Dransfeld, S. Validation of Sentinel-2 LAI products with ground reference data collected across Europe. *Remote Sensing* **11,** 637. `https://doi.org/10.3390/rs11060637` (2019).

10. Caicedo, J. P. R., Verrelst, J., Munoz-Mari, J., Moreno, J. & Camps-Valls, G. Toward a semiautomatic machine learning retrieval of biophysical parameters. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **7,** 1249 – 1259. `https://doi.org/10.1109/JSTARS.2014.2298752` (2014).
11. Calbó, J., Pagès, D. & González, J.-A. Empirical studies of cloud effects on UV radiation: A review. *Reviews of Geophysics* **43,** 1 – 28. `https://doi.org/10.1029/2004RG000155` (2005).
12. Cammarano, D. **andothers**. Use of the Canopy Chlorophyl Content Index (CCCI) for remote estimation of wheat nitrogen content in rainfed environments. *Agronomy Journal* **103,** 1597 – 1603. `https://doi.org/10.2134/agronj2011.0124` (2011).
13. Camps-Valls, G., Tuia, D., Zhu, X. X. & Reichstein, M. Physics-aware machine learning. *Nature Reviews Earth & Environment* **2,** 736–752. `https://doi.org/10.1038/s43017-021-00202-5` (2021).
14. Cao, S. & et al. Learning to normalize: A survey on normalization techniques for deep learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).
15. Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S. & Grégoire, J.-M. Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sensing of Environment* **77,** 22 –33. `https://doi.org/10.1016/S0034-4257(01)00191-2` (2001).
16. Chen, J. M. & Black, T. A. Defining leaf area index for non-flat leaves. *Plant, Cell & Environment* **15,** 421–429 (1992).
17. Chen, J. M., M., R. P., Gower, S. T., Norman, J. M. & Plummer, S. Leaf area index of boreal forests: Theory, techniques, and measurements. *Journal of Geophysical Research: Atmospheres* **102,** 29429 –29443 (1997).
18. Chen, J. **andothers**. TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306* (2021).
19. Chen, L.-C., Papandreou, G., Schroff, F. & Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587.* Published at CVPR 2018 (2017).
20. Clevers, J. G. P. W. The application of a weighted infrared-red vegetation index for estimating leaf area index by correcting for soil moisture. *Remote Sensing of Environment* **29,** 25–37. `https://doi.org/10.1016/0034-4257(89)90076-X` (1988).
21. Copernicus. *Sentinel-2 Documentation* Accessed 10.01.2026. `https://sentiwiki.copernicus.eu/web/document-library#DocumentLibrary-SENTINEL-2DocumentsLibrary-S2-Documents`.
22. Danner, M., Berger, K., Wocher, M., Mauser, W. & Hank, T. Efficient RTM-based training of machine learning regression algorithms to quantify biophysical & biochemical traits of agricultural crops. *ISPRS Journal of Photogrammetry and Remote Sensing* **173,** 278–296 (2021).

23. Daudt, R. C., Le Saux, B. & Boulch, A. Fully convolutional siamese networks for change detection. *IEEE International Conference on Image Processing (ICIP),* 4063–4067. `https://doi.org/10.1109/ICIP.2018.8451652` (2018).
24. De Vries, H. **andothers**. Modulating early visual processing by language. *Advances in Neural Information Processing Systems (NeurIPS),* 6594–6604 (2017).
25. Dey, B. & Ahmed, R. A comprehensive review of AI-driven plant stress monitoring and embedded sensor technology: Agriculture 5.0. *Journal of Industrial Information Integration* **47.** `https://10.1016/j.jii.2025.100931` (2025).
26. Diakogiannis, F. I., Waldner, F., Caccetta, P. & Wu, C. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing* **162,** 94–114 (2020).
27. Disney, M. Terrestrial LiDAR: a three-dimensional revolution in how we look at trees. *New Phytologist* **222.** Includes discussion of canopy structure, anisotropy, and angular effects relevant to optical remote sensing, 1736–1741 (2019).
28. Dumoulin, V., Shlens, J. & Kudlur, M. A learned representation for artistic style. *International Conference on Learning Representations (ICLR)* (2017).
29. Ebel, P., Meraner, A., Schmitt, M. & Zhu, X. X. Multisensor data fusion for cloud removal in global and all-season Sentinel-2 imagery. *IEEE Transactions on Geoscience and Remote Sensing* **59,** 5866–5878. `https://doi.org/10.1109/TGRS.2020.3024744` (2021).
30. Eishoeei, E., Miryaghoubzadeh, M., Erfanian, M., Esfanjani, R. M. & Mancini, M. Soil moisture measurements: a review. *Computers and Electronics in Agriculture* **243.** `https://doi.org/10.1016/j.compag.2025.111379` (2026).
31. European Space Agency. *Sentinel-2 Level-2A Algorithm Theoretical Basis Document (ATBD)* techreport S2-PDGS-MPC-L2A-ATBD-V2.10.0. Available at: ESA Sentinel-2 Processing and Data Quality Ground Segment documentation (ESA, 2021).
32. European Space Agency. *Sentinel-2 Level-2A Product Definition Document* techreport (ESA, 2021).
33. Fan, J. **andothers**. Estimation of potato leaf area index based on spectral information and Haralick textures from UAV hyperspectral images. *Frontiers in Plant Science* **15.** `https://doi.org/10.3389/fpls.2024.1492372` (2024).
34. Fang, H., Baret, F., Plummer, S. & Schaepman-Strub, G. An overview of global leaf area index (LAI): Methods, products, validation, and applications. *Reviews of Geophysics* **57,** 739–799. `https://doi.org/10.1029/2018RG000608` (2019).
35. Fu, X., Wang, J., Huang, Y., Ding, X. & Paisley, J. Pan-sharpening via deep learning: A review. *IEEE Geoscience and Remote Sensing Magazine* **10,** 84–106. `https://doi.org/10.1109/MGRS.2021.3128424` (2022).

36. Galle, N. J. **andothers**. Correlation of WorldView-3 spectral vegetation indices and soil health indicators of individual urban trees with exceptions to topsoil disturbance. *City and Environment Interactions* **11.** `https://doi.org/10.1016/j.cacint.2021.100068` (2021).
37. Ghamisi, P. **andothers**. Multisource and Multitemporal Data Fusion in Remote Sensing. *arXiv preprint.* eprint: `1812.08287` (cs.LG) (2018).
38. Gitelson, A. A. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology* **161,** 165–173. `https://doi.org/10.1078/0176-1617-01176` (2004).
39. Gitelson, A. A., Kaufman, Y. J. & Merzlyak, M. N. Use of a green channel in remote sensing of global vegetation from EOS- MODIS. *Remote Sensing of Environment* **58,** 289 – 298. `https://doi.org/10.1016/S0034-4257(96)00072-7` (1996).
40. Gitelson, A. A. **andothers**. Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophysical Research Letters* **30,** 1248. `https://doi.org/10.1029/2002GL016450` (2003).
41. Guo, A. **andothers**. Inversion of maize leaf area index from UAV hyperspectral and multispectral imagery. *Computers and Electronics in Agriculture* **212,** 108020. `https://doi.org/10.1016/j.compag.2023.108020` (2023).
42. Han, D. **andothers**. Combining Sentinel-1 and -3 imagery for retrievals of regional multitemporal biophysical parameters under a Deep Learning framework. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **15,** 6985 – 6998. `https://doi.org/10.1109/JSTARS.2022.3200735` (2022).
43. He, K., Zhang, X., Ren, S. & Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37,** 1904–1916. `https://doi.org/10.1109/TPAMI.2015.2389824` (2015).
44. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 770–778 (2016).
45. Houborg, R. **andothers**. Joint leaf chlorophyll content and leaf area index retrieval from Landsat data using a regularized model inversion system (REGFLEC). *Remote Sensing of Environment* **159,** 203 – 221. `https://doi.org/10.1016/j.rse.2014.12.008` (2015).
46. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely Connected Convolutional Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 4700–4708 (2017).
47. Huang, Z. & et al. MetaNorm: Learning to Normalize for Domain Generalization. *arXiv preprint arXiv:2203.11111* (2022).

48. Huete, A. **andothers**. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* **83,** 195–213. `https://doi.org/10.1016/S0034-4257(02)00096-2` (2002).
49. Huete, A. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* **25,** 295 – 309. `https://doi.org/10.1016/0034-4257(88)90106-X` (1988).
50. Iglovikov, V. & Shvets, A. TernausNet: U-Net with VGG11 encoder pre-trained on ImageNet for image segmentation. *arXiv preprint arXiv:1801.05746* (2018).
51. ImagineS Project. *ImagineS: Ground Measurements for LAI/FPAR Validation* Ground measurements used as reference for LAI/FPAR product validation. 2018. `https://fp7-imagines.eu`.
52. Ioffe, S. & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International Conference on Machine Learning (ICML),* 448–456 (2015).
53. Jacquemoud, S. & Baret, F. PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment* **34,** 75–91 (1990).
54. Jacquemoud, S. **andothers**. PROSPECT+SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment* **113,** S56–S66. `https://doi.org/10.1016/j.rse.2008.01.026` (2009).
55. Jelinek, Z., Kumhalova, J., Lukas, J. & Miijata, S. Retrieval of Harmonized LAI Product of Agricultural Crops from Landsat OLI and Sentinel-2 MSI Time Series. *Agriculture* **12,** 2080 (2022).
56. Jonckheere, I. **andothers**. Review of methods for in situ leaf area index determination Part I. Theories, sensors and hemispherical photography. *Agricultural and Forest Meteorology* **121,** 19 – 35. `https://doi.org/10.1016/j.agrformet.2003.08.027` (2004).
57. Kamilaris, A. & Prenafeta-Boldú, F. X. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture* **147,** 70–90. `https://doi.org/10.1016/j.compag.2018.02.016` (2018).
58. Kang, Y. **andothers**. A data-driven approach to estimate leaf area index for Landsat images over the contiguous US. *Remote Sensing of Environment* **258.** `https://doi.org/10.1016/j.rse.2021.112383` (2021).
59. Karasiak, N., Dejoux, J.-F., Monteil, C. & Sheeren, D. Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing. *Machine Learning* **111,** 2715 – 2740. `https://doi.org/10.1007/s10994-021-05972-1` (2022).
60. Kattenborn, T., Leitloff, J., Schiefer, F. & Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* **173,** 24–49. `https://doi.org/10.1016/j.isprsjprs.2020.12.010` (2021).
61. Khanal, S., KC, K., Fulton, J. P., Shearer, S. & Ozkan, E. Remote Sensing in Agriculture—Accomplishments, Limitations, and Opportunities. *Remote Sensing* **12.** `https://doi.org/10.3390/rs12223783` (2020).

62. Kirichev, M. M., Slavov, T. S. & Momcheva, G. D. Fuzzy U-Net Neural Network Architecture Optimization for Image Segmentation. *IOP Conference Series: Materials Science and Engineering* **1031,** 012077. `https://doi.org/10.1088/1757-899X/1031/1/012077` (2021).
63. Ko, J., Shin, T., Kang, J., Baek, J. & Sang, W.-G. Combining machine learning and remote sensing-integrated crop modeling for rice and soybean crop simulation. *Frontiers in Plant Science* **15.** `https://doi.org/10.3389/fpls.2024.1320969` (2024).
64. Lai, Y. **andothers**. An empirical model for prediction of wheat yield, using time-integrated Landsat NDVI. *International Journal of Applied Earth Observation and Geoinformation* **72,** 99 – 108. `https://doi.org/10.1016/j.jag.2018.07.013` (2018).
65. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521,** 436–444. `https://doi.org/10.1038/nature14539` (2015).
66. Liu, J. G. Smoothing Filter-based Intensity Modulation: A spectral preserve image fusion technique for improving spatial details. *International Journal of Remote Sensing* **21,** 3461–3472. `https://doi.org/10.1080/014311600750037499` (2000).
67. Liu, Y. **andothers**. The fusion of vegetation indices increases the accuracy of cotton leaf area prediction. *Frontiers in Plant Science* **15,** 1357193 (2024).
68. Loshchilov, I. & Hutter, F. SGDR: Stochastic gradient descent with warm restarts. *International Conference on Learning Representations (ICLR).* arXiv:1608.03983 (2017).
69. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. *International Conference on Learning Representations (ICLR).* arXiv:1711.05101 (2019).
70. Louis, J. **andothers**. Sentinel-2 Sen2Cor: L2A Processor for Users. *Proceedings of the Living Planet Symposium* (2016).
71. Lv, F., Sun, K., Li, W., Miao, S. & Hu, X. Estimation of Leaf Area Index across Biomes and Growth Stages Combining Multiple Vegetation Indices. *Sensors* **24,** 6106 (2024).
72. Ma, T. **andothers**. Development of forest aboveground biomass estimation, its problems and future solutions: A review. *Ecological Indicators* **159.** `https://doi.org/10.1016/j.ecolind.2024.111653` (2024).
73. Maimaitijiang, M. **andothers**. Unmanned Aerial System (UAS)-based phenotyping of soybean using multi-sensor data fusion and extreme learning machine. *ISPRS Journal of Photogrammetry and Remote Sensing* **134,** 43 – 58. `https://doi.org/10.1016/j.isprsjprs.2017.10.011` (2017).
74. Mudò, G. **andothers**. Transgenic expression and activation of PGC-1α protect dopaminergic neurons in the MPTP mouse model of Parkinson's disease. *Cell Mol Life Sci* **69,** 1153–1165 (**october** 2011).
75. Myneni, R. B., Hoffman, S. **andothers**. Global products of vegetation leaf area and fraction absorbed PAR from MODIS. *Remote Sensing of Environment* **83,** 214–231 (2002).

76. Myneni, R. B., Hall, F. G., Sellers, P. J. & Marshak, A. L. Interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing* **33,** 481 – 486. `https://doi.org/10.1109/36.377948` (1995).
77. Myneni, R. B., Ross, J. & Asrar, G. A review on the theory of photon transport in leaf canopies. *Agricultural and Forest Meteorology* **74,** 1–52. `https://doi.org/10.1016/0168-1923(94)02116-H` (1995).
78. NASA. *Landsat 8* `https://science.nasa.gov/mission/landsat-8/`. Accessed 11.01.2026. 2013.
79. NASA. *Landsat 9* `https://science.nasa.gov/mission/landsat-9/`. Accessed 11.01.2026. 2021.
80. ORNL DAAC. *Leaf Area Index Maps at 30-m Resolution, VALERI Site, Larose, Canada* Validation dataset for LAI. 2006. `https://doi.org/10.3334/ORNLDAAC/829`.
81. Paoletti, M. E., Mogollón Gutiérrez, O., Moreno-Avendaño, S., Sancho-Núñez, J. C. & Haut, J. M. A Comprehensive Survey of Imbalance Correction Techniques for Hyperspectral Data Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **16,** 1–19 (2023).
82. Peng, Y., Nguy-Robertson, A., Arkebauer, T. & Gitelson, A. A. Assessment of canopy chlorophyll content retrieval in maize and soybean: Implications of hysteresis on the development of generic algorithms. *Remote Sensing* **9,** 226. `https://doi.org/10.3390/rs9030226` (2017).
83. Perez, E., Strub, F., de Vries, H., Dumoulin, V. & Courville, A. FiLM: Visual Reasoning with a General Conditioning Layer. *Proceedings of the AAAI Conference on Artificial Intelligence* **32** (2018).
84. Radočaj, D., Obhođaš, J., Jurišić, M. & Gašparović, M. Global Open Data Remote Sensing Satellite Missions for Land Monitoring and Conservation: A Review. *Land* **9.** `https://doi.org/10.3390/land9110402` (2020).
85. Reyes-González, A., Kjaersgaard, J., Trooien, T., Hay, C. & Ahiablame, L. Estimation of Crop Evapotranspiration Using Satellite Remote Sensing-Based Vegetation Index. *Advances in Meteorology* **2018.** `https://doi.org/10.1155/2018/4525021` (2018).
86. Richards, J. A. *Remote Sensing Digital Image Analysis: An Introduction* 5th. `https://doi.org/10.1007/978-3-642-30062-2` (Springer, 2013).
87. Roberts, D. R. **andothers**. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40,** 913–929 (2017).
88. Roman, A. & Ursu, T. Multispectral satellite imagery and airborne laser scanning techniques for the detection of archaeological vegetation marks. *Landscape Archaeology on the Northern Frontier of the Roman Empire at Porolissum,* 141–152 (**december** 2016).
89. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI),* 234–241. `https://doi.org/10.1007/978-3-319-24574-4_28` (2015).

90. Rouse, J. W., Haas, R. H., Schell, J. A. & Deering, D. W. Monitoring vegetation systems in the great plains with ERTS. *Environmental Science.* `https://api.semanticscholar.org/CorpusID:133358670` (1973).

91. Rustowicz, R. **andothers**. Semantic segmentation of crop type in Africa: A novel dataset and analysis of deep learning methods. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops,* 75–82 (2019).

92. Schaaf, C. B. **andothers**. First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sensing of Environment* **83,** 135–148. `https://doi.org/10.1016/S0034-4257(02)00091-3` (2002).

93. Shen, W., Li, M., Huang, C. & Wei, A. Quantifying live aboveground biomass and forest disturbance of mountainous natural and plantation forests in Northern Guangdong, China, based on multi-temporal Landsat, PALSAR and field plot data. *Remote Sensing* **8,** 595. `https://doi.org/10.3390/rs8070595` (2016).

94. Sishodia, R. P., Ray, R. L. & Singh, S. K. Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sensing* **12.** `https://doi.org/10.3390/rs12193136` (2020).

95. Szwarcman, D. **andothers**. Prithvi-EO-2.0: A versatile multi-temporal foundation model for Earth observation applications. *arXiv preprint arXiv:2412.02732.* `https://arxiv.org/abs/2412.02732` (2024).

96. Taddeo, S., Dronova, I. & Depsky, N. Spectral vegetation indices of wetland greenness: Responses to vegetation structure, composition, and spatial distribution. *Remote Sensing of Environment* **234.** `https://doi.org/10.1016/j.rse.2019.111467` (2019).

97. Tarrío, J., García, M. & Méndez-Ramos, L. M. Comparison of cloud detection algorithms for Sentinel-2 imagery. *Science of Remote Sensing* **2.** `https://doi.org/10.1016/j.srs.2020.100010` (2020).

98. Tasar, O., Happy, S. L., Tarabalka, Y. & Alliez, P. ColorMapGAN: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks. *IEEE Transactions on Geoscience and Remote Sensing* **58,** 7178–7193. `https://doi.org/10.1109/TGRS.2020.2980417` (2020).

99. Tattaris, M., Reynolds, M. P. & Chapman, S. C. A direct comparison of remote sensing approaches for high-throughput phenotyping in plant breeding. *Frontiers in Plant Science* **7.** `https://doi.org/10.3389/fpls.2016.01131` (2016).

100. Tucker, C. J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* **8,** 127–150. `https://doi.org/10.1016/0034-4257(79)90013-0` (1979).

101. Tuia, D., Persello, C. & Bruzzone, L. Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE Geoscience and Remote Sensing Magazine* **4,** 41 – 57. `https://doi.org/10.1109/MGRS.2016.2548504` (2016).

102. Tziachris, P. & et al. Choosing blocks for spatial cross-validation: lessons from a multi-regional study. *Frontiers in Remote Sensing* (2025).

103. U.S. Geological Survey. *Landsat Collection 2 Level-2 Science Product Guide* Available through the USGS Landsat Science website. 2022.

104. Valanarasu, J. M. J. & Patel, V. M. UNeXt: MLP-based rapid medical image segmentation network. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI),* 23–33 (2022).

105. Van Klompenburg, T., Kassahun, A. & Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture* **177,** 105709. ISSN: 0168-1699. `https://doi.org/10.1016/j.compag.2020.105709` (2020).

106. Verhoef, W. Light scattering by leaf layers with application to canopy reflectance modeling. *Remote Sensing of Environment* **16,** 125–141 (1984).

107. Vermote, E., Justice, C., Claverie, M. & Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment* **185,** 46–56. `https://doi.org/10.1016/j.rse.2016.04.008` (2016).

108. Verrelst, J., Rivera, J. P., Moreno, J. & Camps-Valls, G. Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing* **86,** 157 – 167. `https://doi.org/10.1016/j.isprsjprs.2013.09.012` (2013).

109. Verrelst, J. **andothers**. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. *ISPRS Journal of Photogrammetry and Remote Sensing* **108,** 273–290. `https://doi.org/10.1016/j.isprsjprs.2015.05.005` (2015).

110. Verrelst, J. **andothers**. Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods. *Surveys in Geophysics* **40,** 589–629. `https://doi.org/10.1007/s10712-018-9478-y` (2019).

111. Vivone, G., Dalla Mura, M., Chanussot, J. & Restaino, R. Pansharpening based on convolutional neural networks: A critical review. *IEEE Geoscience and Remote Sensing Magazine* **9,** 56–91. `https://doi.org/10.1109/MGRS.2021.3055011` (2021).

112. Vivone, G. **andothers**. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing* **53,** 2565–2586. `https://doi.org/10.1109/TGRS.2014.2361734` (2015).

113. Vlachopoulos, O. **andothers**. Evaluation of crop health status with UAS multispectral imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **15,** 297 – 308. `https://doi.org/10.1109/JSTARS.2021.3132228` (2022).

114. Wang, L. **andothers**. Scale-aware neural network for semantic segmentation of multi-resolution remote sensing images. *Remote Sensing* **13.** `https://doi.org/10.3390/rs13245015` (2021).

115. Wang, R. **andothers**. A novel semi-empirical model for crop leaf area index retrieval using SAR co- and cross-polarizations. *Remote Sensing of Environment* **296,** 113727. `https://doi.org/10.1016/j.rse.2023.113727` (2023).
116. Wang, Y. & et al. Stacked dilated convolutions and asymmetric architecture for U-Net based medical image segmentation. *Computers in Biology and Medicine* **147,** 105748 (2022).
117. Wang, Z., Chen, J. & Hoi, S. C. H. Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **43,** 3365–3387. `https://doi.org/10.1109/TPAMI.2020.2982166` (2020).
118. Weiss, M. & Baret, F. Evaluation of canopy biophysical variable retrieval performances from the accumulation of large swath satellite data. *Remote Sensing of Environment* **90,** 293–306. `https://doi.org/10.1016/j.rse.2003.12.014` (2004).
119. Weiss, M., Baret, F., Smith, G. J., Jonckheere, I. & Coppin, P. Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling. *Agricultural and Forest Meteorology* **121,** 37–53 (2004).
120. Weiss, M., Jacob, F. & Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment* **236,** 111402. `https://doi.org/10.1016/j.rse.2019.111402` (2020).
121. Xie, Y., Zhang, J., Shen, C. & Xia, Y. TransFuse: Fusing transformers and CNNs for medical image segmentation. *MICCAI* (2022).
122. Yang, Z. **andothers**. An improved scheme for rice phenology estimation based on time-series multispectral HJ-1A/B and polarimetric RADARSAT-2 data. *Remote Sensing of Environment* **195,** 184 – 201. `https://doi.org/10.1016/j.rse.2017.04.016` (2017).
123. Yu, F. & Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122.* Published at ICLR 2016 (2016).
124. Zhang, C., Marzougui, A. & Sankaran, S. High-resolution satellite imagery applications in crop phenotyping: An overview. *Computers and Electronics in Agriculture* **175,** 105584. `https://doi.org/10.1016/j.compag.2020.105584` (2020).
125. Zhang, K. **andothers**. Panchromatic and multispectral image fusion for remote sensing and earth observation: Concepts, taxonomy, literature review, evaluation methodologies and challenges ahead. *Information Fusion* **93,** 227 – 242. `https://doi.org/10.1016/j.inffus.2022.12.026` (2023).
126. Zhang, Z., Liu, Q. & Wang, Y. Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters* **15,** 749–753. `https://doi.org/10.1109/LGRS.2018.2802944` (2018).
127. Zhu, X. X. **andothers**. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine* **5,** 8–36. `https://doi.org/10.1109/MGRS.2017.2762307` (2017).