BULGARIAN ACADEMY OF SCIENCES
INSTITUTE OF MATHEMATICS AND INFORMATICS

ABSTRACT OF DOCTORAL 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"

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

---

Author:
Ventsislav POLIMENOV

Supervisor:
Krassimira IVANOVA
Assoc. Prof. PhD

April 30, 2026

Ventsislav POLIMENOV

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

Supervisor: Assoc. Prof. Krassimira IVANOVA, PhD

The dissertation comprises 138 pages in total, including 20 pages of preliminary (front matter) material, an introduction and five chapters (pp. 1–108), and a bibliography (pp. 109–118). The preliminary section contains the title page, acknowledgements, abstract, table of contents, list of figures, list of tables, and list of abbreviations. The text of the dissertation includes 25 tables and 26 figures. The list of references contains 127 titles.

Chapter 1 Introduction

1.1 Problem Description

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 [11]. Accurate LAI estimation is critical for precision agriculture, enabling monitoring of crop development, optimisation of irrigation and fertilisation, detection of plant stress and prediction of yields. Traditional field-based measurements are labour-intensive and spatially limited [22], while existing satellite-based approaches face their own challenges: empirical vegetation indices saturate at moderate-to-high LAI values, physics-based radiative transfer inversion is computationally demanding and requires site-specific calibration [40], and most critically, current methods process each sensor independently, failing to exploit complementary information across platforms.

Modern Earth observation provides multiple freely available data sources, of which we focus on Sentinel-2 (10 m, 10 bands, 5-day revisit) and Landsat 8/9 (30 m, 6 bands, 8-day revisit), together offering 2-3 day effective revisit. Yet integrating these sensors in a unified framework presents substantial technical obstacles: resolution mismatch, differing spectral configurations, unequal tile counts that can cause mode collapse during training, and ensuring consistent LAI estimates regardless of input sensor. A further challenge, often overlooked in deep learning studies, is spatial data leakage in validation – agricultural landscapes exhibit strong spatial autocorrelation, and random train-test splits inflate performance estimates by placing correlated pixels in both sets [18]. Deep learning offers a potential breakthrough, as convolutional neural networks can learn complex spectral-spatial LAI relationships directly from data, but this introduces the fundamental challenge of obtaining spatially dense training labels when field-measured LAI is unavailable at the required scale.

This thesis addresses these challenges by developing a multi-sensor deep learning framework for operational LAI estimation. The approach combines a U-Net architecture with Conditional Batch Normalisation and Atrous Spatial Pyramid Pooling (ASPP), together with novel training strategies – including batch balancing and VI-ensemble pseudo-labels – to overcome the lack of field-measured ground truth, achieve robust cross-sensor generalisation, and demonstrate strong accuracy in agricultural LAI mapping using Sentinel-2 and Landsat 8/9 imagery.

1.2 Goals and Objectives

The primary goal 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.

The specific objectives are:

1. **Data preprocessing**: Collect and pre-process Sentinel-2 and Landsat 8/9 imagery from pre-corrected surface reflectance products, with consistent tiling across sensors to enable systematic comparative experiments.
2. **Architecture design**: Implement a multi-sensor CNN that processes variable-resolution multispectral data within a unified framework, incorporating Conditional Batch Normalisation for sensor-specific feature adaptation while maintaining shared vegetation representations.
3. **Pseudo-label generation**: Train the model using spatially dense LAI targets generated through an ensemble vegetation index approach, averaging multiple VI-based estimates to reduce individual biases while maintaining sensor-agnostic consistency.
4. **Batch balancing**: Investigate and resolve the epoch-level sensor exposure imbalance (4.2:1 Landsat-to-Sentinel-2 tile ratio) through differential augmentation sampling to prevent mode collapse.
5. **Spatial validation**: Implement spatial block cross-validation using a 4×4 geographic grid to enforce strict separation between training and test sets, preventing spatial data leakage.
6. **Generalisation testing**: Validate cross-region transfer (Plovdiv to Burgas, 200 km) and temporal stability (8-month gap, September 2023 to May 2024) without retraining.
7. **Benchmarking**: Quantify the CNN's value added over constituent vegetation indices and compare against the ESA SNAP Biophysical Processor.
8. **Reproducibility**: Document the complete pipeline, architecture and evaluation methodology to ensure transparency and reproducibility.

1.3 Hypotheses

Three primary hypotheses guide the experimental design:

1. A spectral-aware CNN can generalise across sensors at least as effectively as sensor-specific models, suggesting that learning sensor-invariant LAI features may improve performance compared to sensor-specific

representations.

2. Spatial resolution affects LAI prediction accuracy non-linearly, with the relationship depending on the spatial scale of agricultural features and the strength of landscape autocorrelation.
3. A model trained in one agricultural region can generalise to geographically and temporally distinct regions without retraining, provided predictions are expressed in absolute physical units – cross-region degradation arises primarily from encoding training-region distributional statistics into the output representation rather than from architectural limitations.

1.4 Dissertation Outline

Chapter 2: Research Background reviews satellite remote sensing for LAI estimation, from traditional vegetation indices through radiative transfer models to deep learning, positioning this work within multi-sensor data fusion and biophysical parameter retrieval.

Chapter 3: Methodology details the complete multi-sensor LAI estimation pipeline: VI-ensemble ground truth generation, data loading and tiling at native resolution, quality filtering, spatial block splitting, augmentation and normalisation, the Multi-Sensor U-Net architecture with ASPP and Conditional Batch Normalisation, training strategy with batch balancing, evaluation metrics and the five-tier validation framework.

Chapter 4: Experiments and Results presents spatially rigorous validation of single-sensor baselines and multi-sensor integration (X3: $R^2 = 0.999$, RMSE $= 0.017$ m^2/m^2), cross-region transfer to Burgas (RMSE $= 0.051$ for both sensors), temporal validation over an 8-month gap, comparison with ESA SNAP and 77% RMSE improvement over the best individual VI.

Chapter 5: Conclusions and Future Work synthesises contributions, discusses the circular ground truth limitation transparently, and proposes future directions including field validation campaigns, PROSAIL extension to Landsat and WorldView-3 fine-tuning.

Chapter 2 Research Background

2.1 Literature Review

A central practical focus of this research is the Leaf Area Index (LAI), a key biophysical parameter characterising canopy structure and linked to photosynthesis, respiration and transpiration [1]. LAI is traditionally estimated through labour extensive field sampling or via vegetation indices such as NDVI

and EVI, which offer empirical correlations with canopy properties but suffer from saturation at moderate-to-high LAI and sensor-specific calibration requirements [16]. The increasing availability of high-resolution multispectral imagery from satellites, combined with machine learning techniques, has enabled more scalable and accurate LAI estimation that overcomes these limitations [15, 42]. The present work builds on these developments by implementing a multi-sensor neural network model for LAI estimation using Sentinel-2 and Landsat 8/9 imagery, contributing to the broader effort of developing reliable, non-destructive methods for vegetation monitoring and precision agriculture.

2.2 Satellite Imagery and its Usage for Earth Observation

Remote sensing satellite missions provide multispectral imagery capturing spectral signatures that represent land cover features across the electromagnetic spectrum. The growth of open-data missions has significantly expanded the accessibility of satellite imagery for Earth observation. This section introduces the sensor platforms, spectral properties and biophysical variables relevant to this thesis.

Imagery Types and Resolution

Satellite imagery falls into three principal categories [45]: 1) Panchromatic (PAN) images record a single broad band at high spatial resolution. 2) Multispectral (MS) images record multiple discrete bands – typically visible, NIR and SWIR – exploiting distinctive spectral signatures for vegetation monitoring and land cover classification. 3) Hyperspectral (HS) images capture hundreds of contiguous narrow bands, providing detailed spectral signatures at the cost of large data volumes. The effectiveness of satellite data depends also on spatial resolution (ground area per pixel), spectral resolution (bandwidth of captured signals) and radiometric resolution (bit depth for discriminating energy levels).

Sensor Landscape

The diversity in spatial, temporal and spectral characteristics across platforms (Table 2.1) complicates the application of deep learning to remote sensing, requiring models capable of handling heterogeneous resolutions and data quality [44]. Although the framework developed in this thesis is not limited to any specific sensor pair, it focuses on Sentinel-2 and Landsat 8/9 as the two primary freely available medium-resolution missions with global coverage and operational continuity.

Sentinel-2

Sentinel-2, part of the Copernicus Programme, acquires multispectral imagery across 13 bands – four at 10 m (visible and NIR), six at 20 m (red-edge, SWIR and additional NIR) and three at 60 m (atmospheric correction). The constellation of Sentinel-2A (2015) and Sentinel-2B (2017) achieves a combined five-day revisit.

The red-edge bands at 20 m provide enhanced sensitivity to chlorophyll content and vegetation structure.

TABLE 2.1: Spatial and temporal resolution and spectral differences between satellite platforms. (source: [34, 45])

Satellite	Spatial res. PAN (m)	Spatial res. MS (m)	Spectral bands (#)	Temporal res. (days)	Radiometric res. (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

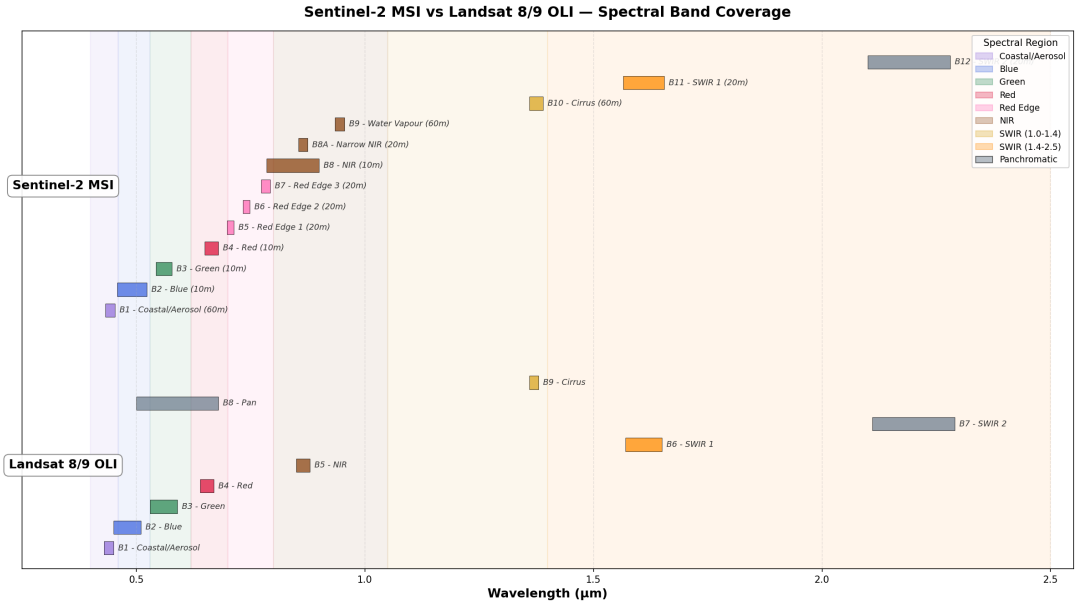

FIGURE 2.1: Spectral response comparison Sentinel-2 and Landsat 8/9.

Landsat 8/9

Landsat 8, operational since 2013, carries the OLI and TIRS sensors with 11 bands at 30 m (VNIR/SWIR) and 15 m (panchromatic), 16-bit radiometric resolution and a 185 km swath in a 16-day sun-synchronous orbit [31]. Landsat 9 (2021) carries improved OLI-2 and TIRS-2 sensors; together the two satellites reduce

the revisit to 8 days [32]. As shown in Figure 2.1, the spectral bands of Sentinel-2 and Landsat 8/9 overlap substantially in the visible, NIR and SWIR regions, while Sentinel-2 provides additional red-edge coverage. This complementarity motivates the multi-sensor fusion approach adopted in this thesis.

Vegetation Indices

Vegetation indices (VIs) combine data from multiple bands into single values that highlight the contrast between vegetation and other surfaces [12, 30]. For LAI estimation, indices incorporating NIR and red-edge bands capture canopy structure more accurately than visible-only indices. This work employs three indices selected for their sensor agnosticism (computable from bands common to both Sentinel-2 and Landsat 8/9) and complementary sensitivity:

NDVI [30] captures overall vegetation vigour through the contrast between NIR and red reflectance (NDVI $= (\rho_{\text{NIR}} - \rho_{\text{Red}})/(\rho_{\text{NIR}} + \rho_{\text{Red}})$), but saturates at LAI $\approx$ 3-4. **GNDVI** [14] substitutes the green band for red, providing greater sensitivity to chlorophyll concentration with a later saturation point (LAI $\approx$ 4-5). **SAVI** [17] introduces a soil brightness correction factor L (SAVI $= (\rho_{\text{NIR}} - \rho_{\text{Red}})(1+L)/(\rho_{\text{NIR}} + \rho_{\text{Red}} + L)$, $L = 0.5$), mitigating soil background influence in sparse canopies.

2.3 Remote Sensing for LAI Estimation: Evolution of Methods

The estimation of LAI from satellite imagery has progressed through a few methodological generations (Figure 2.2), each addressing limitations of its predecessor.

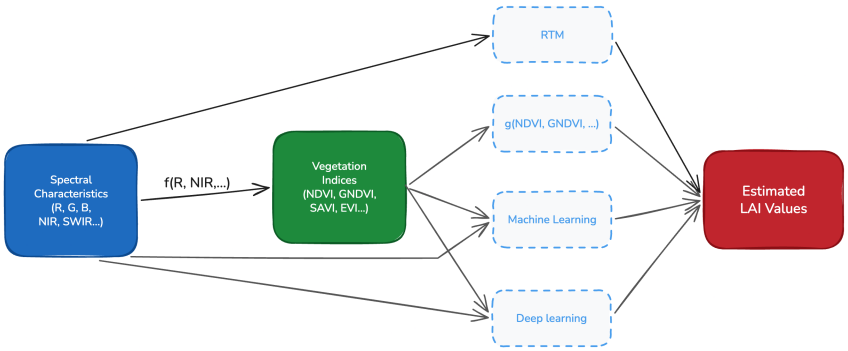

FIGURE 2.2: Schematic overview of methods for deriving Leaf Area Index (LAI) from spectral reflectance measurements.

Physics-Based Radiative Transfer Inversion

Radiative transfer models (RTMs) such as PROSAIL simulate canopy reflectance as a forward function of biophysical parameters and observational geometry

[21]. LAI estimation then becomes an inverse problem: finding the parameter set that minimises the discrepancy between modelled and observed reflectance [40]. However, this inverse problem is ill-posed – multiple parameter combinations can produce nearly identical spectra – requiring strong prior constraints or ensemble sampling methods that are computationally prohibitive for operational processing. Accuracy further depends on correct specification of canopy structure parameters that vary spatially and are rarely known *a priori*.

Vegetation Indices

Rather than inverting a full physical model, vegetation indices (VIs) exploit the known spectral response of green vegetation through algebraic combinations of reflectance bands.

VIs serve a dual role. First, as biophysical proxies: calibrated transfer functions (typically exponential, grounded in Beer–Lambert law) convert index values directly to LAI estimates, e.g. $\text{LAI} = -\frac{1}{k}\ln(1 - f(\text{VI}))$ [6]. Second, as engineered features for downstream statistical and machine learning models. Despite their utility, all empirical indices share fundamental limitations: saturation at moderate-to-high LAI (NDVI plateaus around LAI = 3–5), sensor-specific calibration requirements and inability to represent complex non-linear spectral–LAI relationships across diverse landscapes [18].

Statistical and Machine Learning Regression

Machine learning methods bypass explicit physical modelling by learning empirical mappings from spectral features to LAI through supervised training [41]. Shallow models like Random Forests, Support Vector Regression, Gaussian processes improved upon single-index calibration by combining information from multiple indices and spectral bands simultaneously, partially overcoming individual saturation limitations. However, these methods still rely on hand-crafted features and treat each pixel independently, failing to exploit the spatial structure inherent in multispectral imagery.

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

Deep convolutional neural networks represent a fundamental shift – rather than relying on hand-crafted vegetation indices, CNNs learn hierarchical feature representations directly from raw spectral bands through end-to-end optimisation [23, 47]. For LAI estimation, CNNs offer learning of complex non-linear spectral-spatial relationships without saturation, exploitation of spatial context at multiple scales, end-to-end optimisation without intermediate feature engineering and potential for transfer learning across sensors and regions [43]. The critical bottleneck is the requirement for spatially dense training labels, which this thesis addresses through an ensemble vegetation index approach generating pseudo-labels for supervised training (Section 3.1).

2.4 Deep Learning Architectures for Dense LAI Prediction

Dense biophysical parameter estimation from satellite imagery constitutes a semantic segmentation task – assigning a continuous LAI value to every pixel. This section reviews the architectural components employed in this thesis.

U-Net: Encoder-Decoder Architecture

The U-Net architecture [35] employs an encoder that progressively downsamples through convolutional and pooling layers, capturing increasingly abstract features, and a symmetric decoder that recovers spatial resolution through transposed convolutions. The critical innovation is skip connections, a direct concatenation of encoder and decoder features at matching resolutions, which reintroduce fine spatial detail lost during downsampling, provide alternative gradient pathways mitigating vanishing gradients and fuse coarse semantic information with fine-grained spatial detail at each resolution level.

These properties align well with satellite-based LAI estimation. Agricultural LAI exhibits structure across multiple scales, from crop rows to field-scale zones to landscape gradients; the encoder captures this hierarchy while skip connections preserve fine detail. The fully convolutional design processes images of arbitrary size, enabling training on small tiles with inference on full scenes [24]. U-Net variants have been widely adopted across remote sensing – road extraction [46], building detection [19], crop classification [36], change detection [7], canopy height modelling [24] and biomass estimation [37]. For LAI specifically, U-Net has demonstrated 30-50% RMSE improvements over VI baselines [33, 43], though most prior work focuses on single-sensor scenarios.

Dilated Convolutions and ASPP

Expanding the receptive field is critical for capturing large-scale context. Dilated (atrous) convolutions [5] insert $(r-1)$ zeros between kernel weights:

$$y[i] = \sum_k w[k] \cdot x[i + r \cdot k] \tag{2.1}$$

A 3×3 kernel with rate r=2 covers a 5×5 region with only 9 parameters, preserving spatial resolution while efficiently expanding context. Atrous Spatial Pyramid Pooling (ASPP) [5] applies multiple dilated convolutions in parallel at different rates (e.g. $r \in \{1, 6, 12, 18\}$), concatenates the outputs and fuses via 1×1 convolution, optionally including global average pooling for image-level context. Integrated at the U-Net bottleneck, ASPP enables simultaneous reasoning across sub-field, field and landscape scales [9].

Batch Normalisation

Batch Normalisation (BN) [20] normalises layer activations to zero mean and unit variance within each mini-batch:

$$\mathrm{BN}(x) = \gamma \frac{x - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} + \beta \tag{2.2}$$

where γ, β are learnable scale and shift parameters. BN prevents gradient explosion, enables higher learning rates and provides implicit regularisation [20]. During inference, exponential moving averages of training statistics replace batch statistics.

Conditional Batch Normalisation for Multi-Sensor Learning

In multi-sensor settings, standard BN computes statistics across all samples, forcing a compromise representation that averages over sensor-specific distributional differences [10]. Conditional BN [8, 10] addresses this by conditioning on sensor identity, learning sensor-specific parameters $(\gamma^{(s)}, \beta^{(s)})$:

$$\mathrm{CondBN}(x, s) = \gamma^{(s)} \frac{x - \mu_{\mathcal{B}_s}}{\sqrt{\sigma_{\mathcal{B}_s}^2 + \epsilon}} + \beta^{(s)} \tag{2.3}$$

where $s \in \{\mathrm{S2}, \mathrm{L8}\}$ and batch statistics are computed only from sensor s samples. This maintains sensor-specific normalisation while sharing all convolutional weights, adding under 1% to total parameters. The core insight – affine parameters encode domain-specific statistics while convolutional weights encode shared semantic content – originates from neural style transfer [10] and applies directly to multi-sensor fusion, where "style" becomes sensor-specific radiometric characteristics and "content" is the underlying LAI pattern. Conditional BN requires single-sensor batches for valid statistics computation.

Foundation Models for Earth Observation

Large-scale foundation models pre-trained on massive satellite archives represent an emerging paradigm. Prithvi-EO-2.0 [38], developed by IBM and NASA, is a Vision Transformer with masked autoencoder pre-training on 4.2 million HLS image sequences, available in configurations up to 600M parameters. While demonstrating strong performance on segmentation tasks, Prithvi has not been evaluated for LAI retrieval. The two approaches are complementary: foundation model pre-training could provide improved feature representations, while our multi-sensor training strategy and spatial validation methodology remain essential for operational LAI retrieval. Our task-specific U-Net with conditional BN achieves remarkable results with only 28.7M parameters.

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.

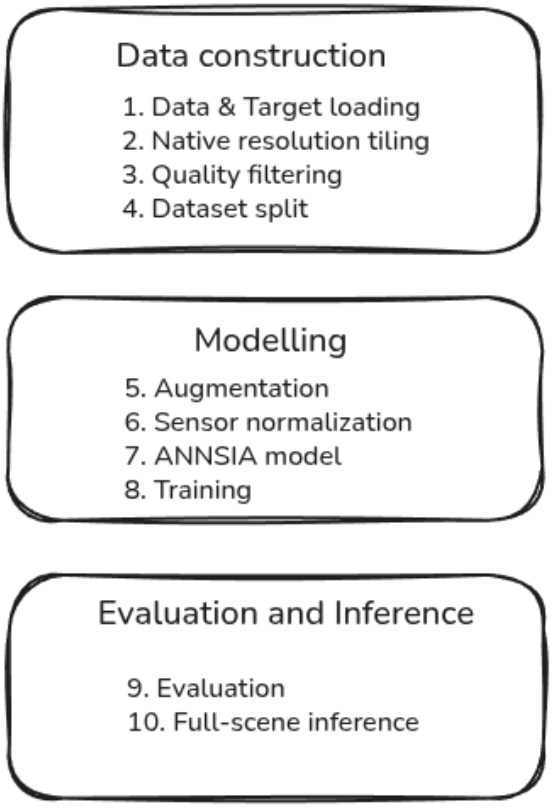

FIGURE 3.1: Overview of the pipeline for multi-sensor LAI estimation, showing the sequence from pseudo-label generation to model training and validation.

3.1 LAI Ground Truth Generation

Supervised deep learning for LAI estimation requires large volumes of labelled training data, yet field-based LAI measurements are destructive, labour-intensive and spatially sparse – none were available for the Plovdiv study region during the 2023 growing season. Existing operational products present their own limitations: ESA SNAP's Biophysical Processor tends to extrapolate beyond VI

saturation limits in dense canopies, while the Copernicus Global Land Service (CGLS) operates at 300 m resolution. This work instead adopts **ensemble vegetation index-based pseudo-labels** as proxy training targets, positioning them as physically grounded, reproducible proxies that enable spatial pattern learning while maintaining transparency about their limitations.

Vegetation Index Selection

Three indices were selected based on sensor agnosticism (computable from bands common to both Sentinel-2 and Landsat 8/9), complementary sensitivity and established empirical validation [25, 29]. NDVI captures overall vegetation vigour and canopy density, GNDVI enhances sensitivity to chlorophyll concentration with a later saturation point and SAVI mitigates soil background effects in sparse canopies. Table 3.1 summarises the formulas and band mappings.

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

VI-to-LAI Transfer Functions

All three indices use the exponential (Beer-Lambert) transfer function, which models LAI as the optical thickness required to attenuate incident radiation:

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

where VI is clipped to $[0, 0.95]$ and $k = 0.60$ is the empirical extinction coefficient, adopted from established agricultural calibrations [6, 13, 17]. The Beer-Lambert formulation has strong theoretical grounding in radiative transfer theory, as the attenuation 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.

Ensemble Aggregation

To leverage complementary strengths and reduce method-specific biases, ensemble averaging is employed:

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

where $N = 3$. Inter-method variability provides pixel-wise uncertainty:

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

High inter-method standard deviation ($\sigma > 0.5$ m^2/m^2) indicates saturation onset, soil or atmospheric contamination, or mixed pixels. The uncertainty metric is stored alongside the ensemble mean and used for training sample quality filtering (excluding tiles with $\sigma > 1.0$ m^2/m^2) and for identifying spatial regions where estimates are less reliable. Ensemble averaging partially cancels random errors, extends the effective dynamic range through differing saturation thresholds and reduces sensitivity to single-method failures. A CNN trained on ensemble targets can subsequently outperform individual indices by learning spatial patterns beyond pixel-level spectral relationships, as demonstrated in Chapter 4.2.

Justification and Limitations

Training a CNN on VI-derived labels introduces a fundamental circularity where the model learns to spatially interpolate and smooth VI-ensemble LAI rather than retrieving absolute LAI independently. This work does not claim absolute LAI retrievals; it is positioned as **operational LAI estimation using VI-ensemble pseudo-labels**. Several factors support this approach: the ensemble aggregates three independent formulations with decades of empirical validation [13, 17, 39]; the Beer–Lambert function is grounded in radiative transfer theory; ESA validation reports RMSE < 1.0 m^2/m^2 for VI-based approaches in agricultural systems [3]; and the labels are derived independently of SNAP, preserving it as an orthogonal validation reference (Section 4.4).

The principal limitations are threefold. First, the circular ground truth means the CNN's task is spatial interpolation of VI-ensemble LAI; this is addressed through complementary assessments including comparison against SNAP, spatial coherence analysis, and cross-region generalisation experiments (Chapter 4). Second, VI saturation at LAI 3–5 m^2/m^2 is inevitably inherited by the CNN, though this is mitigated by focusing on agricultural landscapes where LAI typically remains below 5 m^2/m^2. Third, training data derive from a single season (September 2023), though temporal and cross-region generalisation are tested using May 2024 data from Burgas (Section 4.3).

3.2 Data Pipeline, Training and Inference

This section describes the end-to-end processing pipeline, from satellite data ingestion through model training to full-scene LAI prediction.

Data Loading and Sensor Configuration

The methodology employs Level-2 atmospherically corrected surface reflectance products: Sentinel-2 Level-2 from the Copernicus Data Space and Landsat 8/9 Collection 2 Level-2 from USGS Earth Explorer. For Sentinel-2, all bands at 20 m resolution or better are retained (10 bands total), excluding the 60 m coastal, aerosol and cirrus bands (B01, B09, B10). For Landsat 8/9, all six 30 m VNIR–SWIR bands (B2–B7) are used. Six common spectral regions (Blue, Green, Red, NIR, SWIR1, SWIR2) form a sensor-agnostic subset, while Sentinel-2's four red-edge bands provide additional spectral detail. Table 3.2 summarises the band correspondence. Raw digital numbers are converted to physical reflectance values (dividing by 10 000 for Sentinel-2; applying USGS gain/offset for Landsat) and stored as `float16` for efficient storage.

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**	

Tiling at Native Resolution

Unlike conventional approaches that resample all inputs to a common grid, the pipeline preserves native sensor resolution throughout. Each sensor is tiled at its original pixel spacing, with tile dimensions adjusted to cover approximately the same geographic area (Table 3.3). This avoids resampling artifacts that degrade fine-scale spatial information and respects each sensor's point spread function. The CNN architecture accommodates different resolutions through sensor-specific pathways (Section 3.2). Full scenes are subdivided into non-overlapping tiles;

partial tiles at scene boundaries are discarded if below 90% of nominal dimensions. Each tile retains its coordinate reference system and geotransform metadata.

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

Quality Filtering and Masking

Both sensors provide pixel-level quality layers for cloud and artifact masking. For Sentinel-2, the Sen2Cor Scene Classification Layer (SCL) is used to mask clouds, cloud shadows, cirrus, snow, saturated/defective and no-data pixels [28]. For Landsat 8/9, conservative masking is applied via bit-packed QA_PIXEL flags for high-confidence cloud, shadow, cirrus, saturated and fill pixels. After pixel-level masking, tiles are retained only if they meet three criteria: valid fraction ≥ 0.95, LAI ensemble mean $\in [0, 10]$ m^2/m^2, and LAI ensemble standard deviation < 1.5 m^2/m^2. Masked pixels within retained tiles are set to NaN and excluded from loss computation.

Dataset Generation and Spatial Splitting

Processed tiles are serialised as `.npy` files containing reflectance arrays (10 channels for Sentinel-2, 6 for Landsat), ensemble mean LAI and its uncertainty, all in `float16`. The final experiments (Exp X1-X4, Chapter 4.1) implement geographic disjointness through a 4×4 spatial grid partitioning (Figure 4.1). Blocks are randomly assigned to training, validation and test splits using a fixed seed with a 70/15/15 ratio. Across both Plovdiv scenes this yields 22 non-empty blocks (15 training, 3 validation, 4 testing). This block-level assignment ensures that all tiles within the same geographic region are kept in the same split, preventing spatial data leakage.

Data Augmentation

Training data are augmented on-the-fly using eight geometric transformations: the identity, three 90° rotations and four flip/rotation combinations. Only 90° increments and axis-aligned flips are used, preserving spectral values exactly without interpolation. Spectral augmentations were deliberately omitted to maintain the physical integrity of spectral-LAI relationships. For multi-sensor experiments, Sentinel-2 tiles are sampled with a fourfold higher augmentation rate to compensate the ∼4:1 Landsat tile count imbalance, equalising sensor exposure at the batch level.

Normalisation

Sensor-specific Z-score normalisation is applied to map each sensor's reflectance distribution to a common standardised scale:

$$x_{\text{norm}}^{(s,b)} = \frac{x_{\text{raw}}^{(s,b)} - \mu^{(s,b)}}{\sigma^{(s,b)}} \tag{3.4}$$

where $\mu^{(s,b)}$ and $\sigma^{(s,b)}$ are the mean and standard deviation for sensor s and band b, computed from the training split only and applied consistently to validation and test sets. This linear transformation preserves relative spectral contrasts critical for vegetation index computation while facilitating shared representation learning despite inter-sensor differences.

Model Architecture

The core architecture is ANNSIA – a multi-sensor U-Net with Conditional Batch Normalisation and Atrous Spatial Pyramid Pooling (ASPP), illustrated in Figure 3.2.

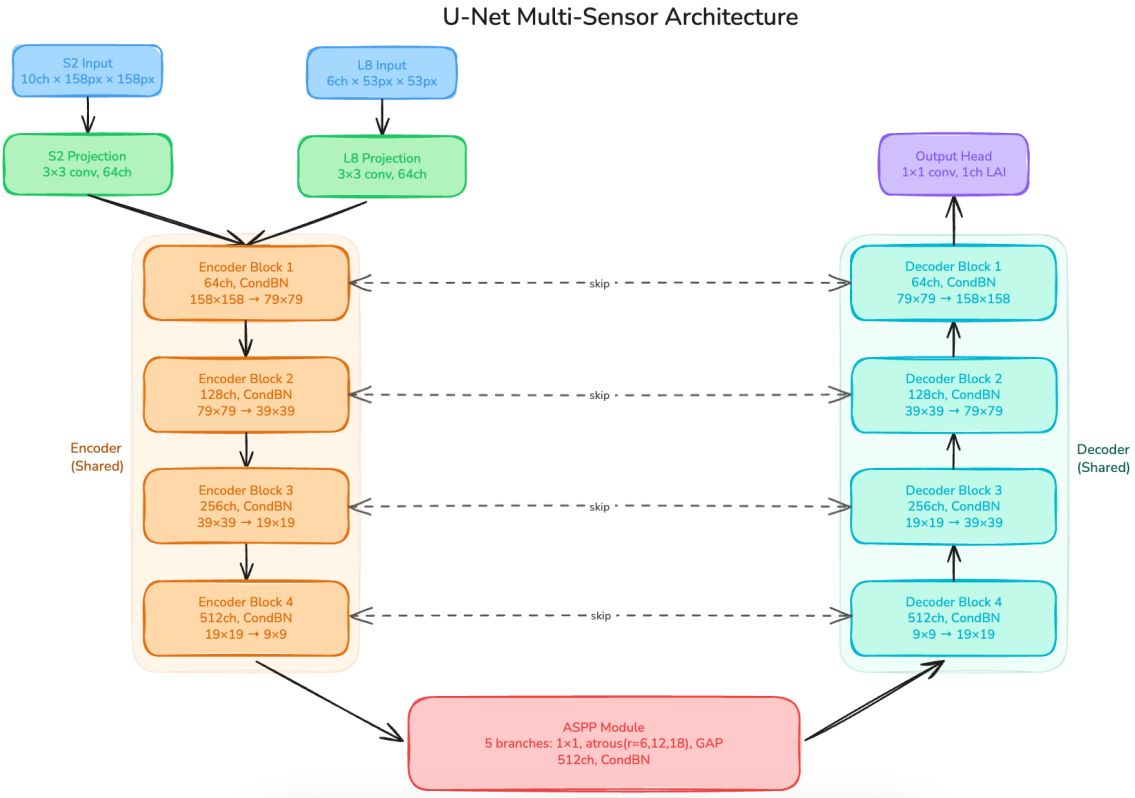

FIGURE 3.2: The architecture of ANNSIA.

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.

Sensor-specific 3×3 input projections map the different channel counts (10 for Sentinel-2, 6 for Landsat) to a common 64-channel representation. The shared encoder consists of four downsampling blocks (channels: $64 \to 128 \to 256 \to 512$), each comprising two 3×3 convolutions with ReLU activation, Conditional Batch Normalisation and 2×2 max pooling. At the bottleneck, an ASPP module [5] aggregates multi-scale context through five parallel branches: a 1×1 convolution, three 3×3 atrous convolutions at rates 6, 12 and 18 and global average pooling – capturing sub-field (∼200 m), field (∼400 m) and multi-field (∼600 m) context respectively. The symmetric decoder recovers spatial resolution through four upsampling blocks with transposed convolutions and skip connections. A final 1×1 convolution with ReLU activation produces the single-channel LAI output. The network contains approximately 28.7 million trainable parameters.

Conditional Batch Normalisation

Although inputs are Z-score normalised, Sentinel-2 and Landsat 8/9 exhibit different internal feature distributions due to differing input resolutions, spectral content and spatial frequency characteristics. Conditional Batch Normalisation addresses this by maintaining sensor-specific affine parameters $(\gamma^{(s)}, \beta^{(s)})$ and running statistics at every normalisation layer, while sharing all convolutional weights. During training, each batch contains samples from a single sensor; during inference, the sensor identifier selects the appropriate parameters. This enables sensor-invariant convolutional filters with sensor-specific feature scaling.

Training Strategy

Training uses the AdamW optimiser [27] with learning rate $\eta = 10^{-3}$, weight decay $\lambda = 10^{-5}$ and cosine annealing with warm restarts [26]:

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min}) \left(1 + \cos\left(\frac{T_{\text{cur}}}{T_i}\pi\right)\right) \tag{3.5}$$

where $\eta_{\max} = 10^{-3}$, $\eta_{\min} = 10^{-6}$ and T_i is the restart period. The loss function is NaN-aware pixel-wise MSE over valid pixels:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{|\mathcal{V}|} \sum_{(i,j)\in\mathcal{V}} \left(\hat{y}_{i,j} - y_{i,j}\right)^2 \tag{3.6}$$

where $\mathcal{V}$ is the set of non-NaN pixels after cloud masking. A batch size of 64 per sensor is used on an NVIDIA A100 40 GB GPU (Section 4.5), with single-sensor batches alternating between Sentinel-2 and Landsat 8/9 to ensure correct operation of Conditional Batch Normalisation.

Evaluation Metrics

Model performance is quantified using RMSE, MAE, R^2, Bias and Variance Ratio. In all equations, $\hat{y}_i$ denotes the predicted LAI, y_i the reference LAI and N the number of valid pixels:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{3.7}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{3.8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{3.9}$$

$$\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i) \qquad \text{Variance Ratio} = \frac{\sigma_{\hat{y}}}{\sigma_y} \tag{3.10}$$

Metrics are computed per sensor, stratified by LAI bin (0-0.5, 0.5-1, 1-2, 2-3, 3-5, 5-10 m^2/m^2) and per scene. Band importance is assessed via permutation – shuffling band b and measuring the RMSE increase relative to baseline.

Inference

For operational LAI mapping, trained models are applied to full scenes using a sliding window with ∼25% overlap (40 pixels for Sentinel-2, 13 for Landsat). Within overlap regions, predictions are blended via linear-ramp weighting:

$$\hat{y}_{i,j} = \frac{\sum_k w_k(i,j)\,\hat{y}_k(i,j)}{\sum_k w_k(i,j)} \tag{3.11}$$

where w_k ramps linearly from 0 to 1 over $O/2$ pixels at each tile edge, eliminating boundary discontinuities. Output LAI maps are exported as compressed GeoTIFF rasters with the original coordinate reference system. Full-scene inference for a Sentinel-2 granule (∼120 million pixels) completes in under two minutes on a single A100 GPU.

3.3 Validation Methodology

Using vegetation index-derived ensemble estimates as training targets introduces a fundamental circularity where the CNN learns to predict LAI from reflectance using labels 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. This work employs a five-tier validation framework, explained below. No

single tier overcomes the circular ground truth limitation, but convergence across multiple independent axes increases confidence in the methodology's robustness.

Tier 1: Internal Machine Learning Validation

Traditional random splitting introduces spatial data leakage, as neighbouring tiles exhibit strong spatial autocorrelation from shared land cover, phenology and atmospheric conditions. To enforce geographic disjointness, a 4×4 spatial block partitioning strategy (Section 3.2) divides each scene into 16 equal blocks, randomly assigned with approximately 70% training, 15% validation and 15% testing. Across the two Plovdiv scenes this yields 22 non-empty blocks (15 training, 3 validation, 4 testing), with block-level assignment ensuring that spatially proximate tiles remain in the same partition. Model performance is assessed using RMSE, MAE, R^2, bias and variance ratio (formulations in Section 3.2), stratified into four LAI bins – low (0-2), medium (2-4), high (4-6) and very high (>6 m^2/m^2) – to diagnose saturation effects and LAI-dependent biases.

Tier 2: Operational LAI Products

The SNAP Biophysical Processor implements neural network emulation of PROSAIL [21], trained on 50,000 simulated canopy configurations. It provides pixel-level LAI estimates grounded in radiative transfer theory rather than empirical indices and is not used as training data at any stage, preserving its independence. Agreement between CNN predictions and SNAP indicates consistency with physics-based retrievals; disagreement highlights divergence between VI-ensemble pseudo-labels and RTM approaches.

Tier 3: Radiative Transfer Model Validation

The PROSAIL model [21] couples the PROSPECT leaf optical properties model with the SAIL canopy reflectance model. To test consistency with radiative transfer theory, PROSAIL is run for selected fields with known crop types using literature-based parameter ranges and measured solar geometry. The simulated reflectance spectra are passed to the trained CNN and the resulting predictions compared against the forward-simulation LAI values. Consistency (RMSE below 0.5 m^2/m^2) indicates that the learned spectral-LAI relationship aligns with physics-based theory. This tier is limited by PROSAIL parameter uncertainty and should be interpreted as a physics-based sanity check independent of empirical VIs.

Tier 4: Cross-Region and Temporal Consistency

The model trained on Plovdiv September 2023 scenes is evaluated on an independent site – the Burgas region, approximately 200 km east – across two seasons. Geographic transfer is assessed through zero-shot application to Burgas autumn 2023 scenes, isolating geographic displacement while controlling for phenology. Temporal transfer is assessed through zero-shot application to Burgas

spring 2024 scenes, introducing an eight-month seasonal gap spanning an entire phenological cycle. A degradation exceeding 20% RMSE increase would indicate overfitting to Plovdiv-specific patterns. Multi-sensor training further enables internal consistency checks: CNN predictions from Sentinel-2 and Landsat 8 over spatially overlapping areas should agree within ± 0.5 m^2/m^2.

Tier 5: Empirical Vegetation Index Benchmarking

The most direct measure of the CNN's value-added contribution is comparison against the individual VI methods (NDVI-exp, GNDVI-exp, SAVI-exp) and their ensemble mean used to generate training labels. If the CNN merely memorises 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, LAI estimates from all three VI methods, the ensemble mean and the CNN are compared on RMSE, MAE and R^2.

Rationale and Limitations

Convergent evidence across five independent axes provides confidence despite the circular ground truth limitation. Tier 1 confirms out-of-sample generalisation, Tier 2 contextualises predictions within operational products, Tier 3 verifies consistency with radiative transfer theory, Tier 4 demonstrates geographic & temporal transferability and Tier 5 quantifies value added beyond pixel-level indices. This framework does not claim to validate absolute LAI accuracy against field measurements; the work is positioned as operational LAI estimation with demonstrated spatial learning, multi-sensor fusion and generalisation capacity. A future sixth tier incorporating field LAI measurements would enable absolute accuracy assessment and crop-specific calibration.

Chapter 4 Experiments and Results

4.1 Experimental Configuration

This section describes the dataset, spatial splitting strategy and model variants evaluated in this thesis.

Study Region and Data Acquisition

The primary dataset comprises satellite imagery acquired over the **Plovdiv region, Bulgaria** (42.15°N, 24.75°E), a predominantly agricultural landscape characterised by cereal cultivation (wheat, maize), industrial crops (sunflower) and viticulture, with flat to gently rolling topography (100–300 m elevation). All training data were collected during September 2023, corresponding to late-season phenology that provides a diverse LAI range (0–6 m^2/m^2) while avoiding extreme saturation conditions. Generalisation is assessed using independent scenes from the Burgas

region, approximately 200 km east. Table 4.1 summarises the satellite scenes used.

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.

The Sentinel-2 and Landsat 9 training acquisitions are separated by only one day, minimising phenological change and enabling consistent VI-ensemble ground truth generation. The Burgas site provides a fully independent test with different agricultural composition, soil characteristics and coastal influences, testing both geographic transfer (200 km) and temporal transfer (8-month gap) simultaneously.

Tile Counts and Data Distribution

After tiling, quality filtering and spatial block splitting, the dataset comprises 17,519 tiles. Landsat 9 tiles outnumber Sentinel-2 by approximately 4:1 due to the coarser resolution producing more tiles per scene, motivating the batch balancing strategy (Section 3.2).

Normalisation Statistics

Table 4.2 presents per-band mean and standard deviation for the six spectral regions common to both sensors. Landsat 9 reflectances are systematically lower in the visible bands, reflecting differences in spectral response functions and atmospheric correction, while NIR and SWIR show closer agreement. These sensor-specific statistics are applied via Z-score normalisation (Equation (3.4)) and stored with the model checkpoint for consistent use during validation and inference.

TABLE 4.2: Per-band normalisation statistics for 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

LAI Ground Truth Statistics

The VI-ensemble LAI distribution (Table 4.3) has a mean of 1.82 m^2/m^2 and is right-skewed, reflecting the prevalence of low-to-moderate LAI with occasional high-LAI fields. Training, validation and test sets exhibit nearly identical statistics (means within 0.03 m^2/m^2), confirming that the spatial block split preserves LAI diversity. The maximum observed LAI of 5.92 m^2/m^2 remains below the extreme saturation threshold. Mean inter-VI standard deviation is 0.34 m^2/m^2 (CV = 19%), indicating reasonable consensus among the three estimation methods.

TABLE 4.3: LAI ensemble statistics across splits (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**

Spatial Block Split

All experiments employ spatial block cross-validation. Each scene is divided into a 4×4 grid of equal geographic blocks, randomly assigned to training, validation and test splits using a fixed seed with a 70/15/15 ratio. Across both Plovdiv scenes this yields 22 non-empty blocks (15 training, 3 validation, 4 testing). This block-level assignment ensures that all tiles within the same geographic region remain in the same split, preventing spatial data leakage from neighbouring tiles appearing in different partitions (Figures 4.1 and 4.2).

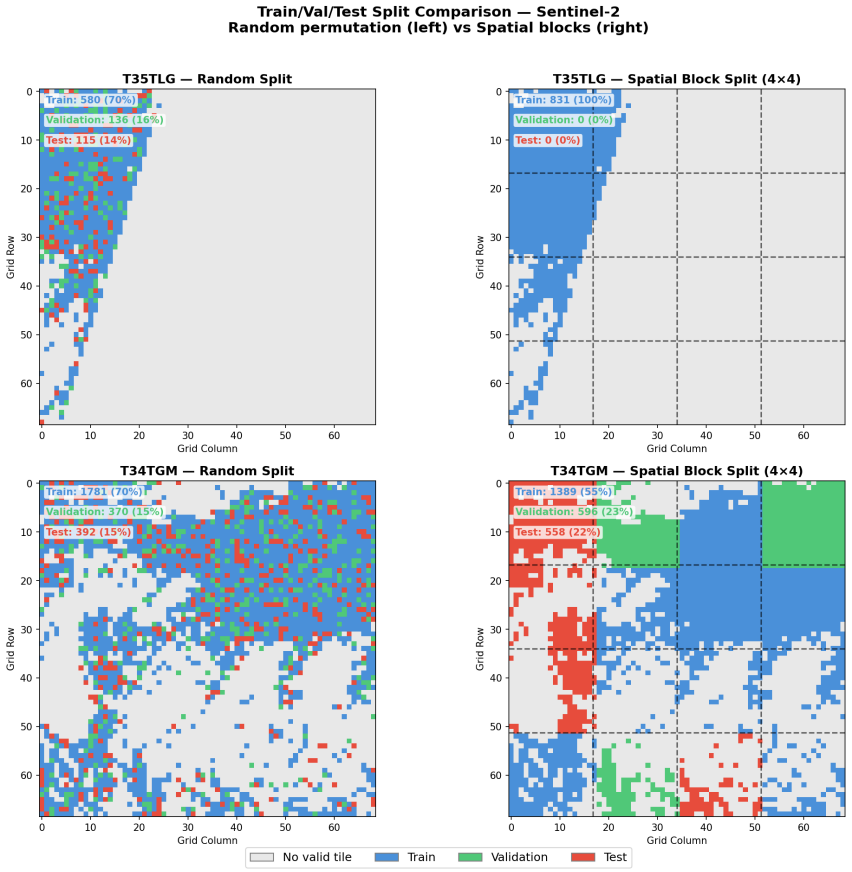

FIGURE 4.1: Random split versus spatial block split for Sentinel-2 images T35TLG and T34TGM.

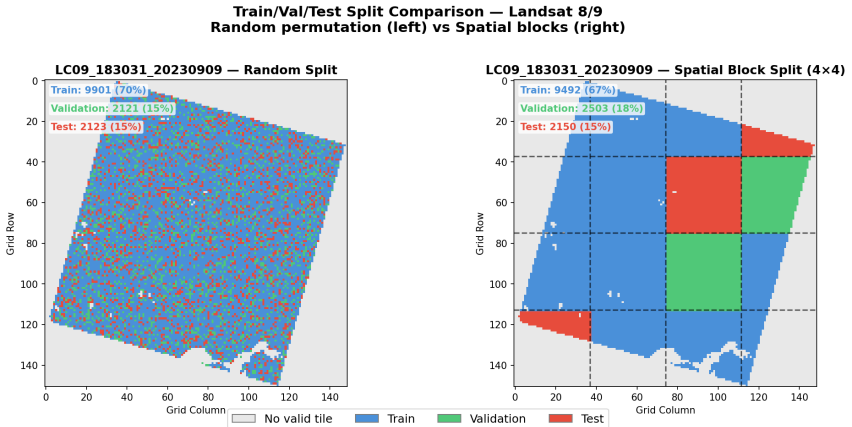

FIGURE 4.2: Random split versus spatial block split for Landsat 9 image LC09_183031.

Model Variants

Four experimental configurations isolate the contributions of multi-sensor integration, ground truth formulation and normalisation. **X1** (Sentinel-2 only, 10 bands at 10 m) and **X2** (Landsat 8/9 only, 6 bands at 30 m) establish single-sensor

baselines. **X3** (multi-sensor, VI-ensemble targets, no LAI normalisation, 4× batch balancing) is the **best-performing model** and reference for all generalisation experiments. **X4** (identical to X3 but with hybrid VI+PROSAIL targets at 50:50 ratio) ablates whether physics-based RTM constraints improve generalisation. Table 4.4 provides a summary of experiments and key results across the validation framework.

TABLE 4.4: Summary of experiments and key results across the validation framework.

#	Tier	Experiment	Focus	Primary Metrics	Key Results
Single-Sensor Baselines (Section 4.2)					
X1	T1	Sentinel-2 Only	Reference at 10 m, 10 bands	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, No LAI Norm (**best**)	Joint S2 + L8/9	Test RMSE, R^2	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	RTM targets vs. VI-only	ΔRMSE vs. X3	+35% RMSE degradation
Generalisation (Section 4.3)					
G1	T4	Cross-Region (Burgas, 200 km)	Zero-shot geographic transfer	RMSE	Both sensors: RMSE = 0.051
G2	T4	Temporal Transfer (8-month gap)	Phenological consistency	LAI ratio	S2: 2.9×; L8: 1.7× increase
G3	T5	CNN vs. Vegetation Indices	Value added beyond VIs	RMSE vs. VI-ensemble	CNN: 0.010 (77% lower than best VI)
Operational Validation (Section 4.4)					
O1	T1, T4	Full-Scene Inference	Deployment-scale consistency	RMSE, VR	S2: 0.004, VR = 0.999; L8: 0.010, VR = 0.998
O2	T2	ESA SNAP Comparison	Operational RTM product	RMSE, R^2	SNAP vs. VI-ens.: RMSE = 1.987, R^2 = 0.312

Note: All RMSE values in m^2/m^2. Validation tiers: T1 = Internal ML, T2 = Operational products, T3 = RTM, T4 = Cross-sensor/temporal, T5 = VI benchmarking. VR = Variance Ratio.

4.2 Spatially Rigorous Validation

This section presents the core performance evaluation of all model variants under the spatial block validation framework (Section 4.1). All metrics reflect genuine out-of-region prediction accuracy, with strict geographic disjointness between training and test sets.

Single-Sensor Baselines

Single-sensor U-Net models were trained under identical spatial validation constraints to isolate the contribution of multi-sensor fusion. Both use the same encoder-decoder architecture adapted for sensor-specific input dimensions (10 channels for Sentinel-2, 6 for Landsat 8/9). Table 4.5 summarises test set performance.

TABLE 4.5: Single-sensor baseline performance under spatial block validation (test set).

Exp.	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	$+0.0030$

Sentinel-2 achieves lower RMSE (0.059 vs. 0.071 m^2/m^2) owing to finer spatial resolution and the red-edge bands (705, 740, 783 nm) that provide direct sensitivity to chlorophyll content. Despite coarser resolution and the absence of red-edge bands, Landsat achieves excellent $R^2 = 0.990$. Its higher R^2 despite higher RMSE reflects a resolution-dependent variance effect: 30 m pixels inherently smooth within-field variability, reducing scatter around the regression line. Both models exhibit negligible bias (<0.003 m^2/m^2), confirming that learned spectral-LAI relationships are not spatial memorisation artifacts.

Experiment X3: Best Model Performance

The multi-sensor model combines Sentinel-2 and Landsat 8/9 through Conditional Batch Normalisation with VI-ensemble pseudo-labels, no LAI target normalisation and $4\times$ Sentinel-2 batch balancing. Table 4.6 and Figure 4.3 present the test set results.

TABLE 4.6: Experiment X3 test set performance: multi-sensor model without LAI normalisation.

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**

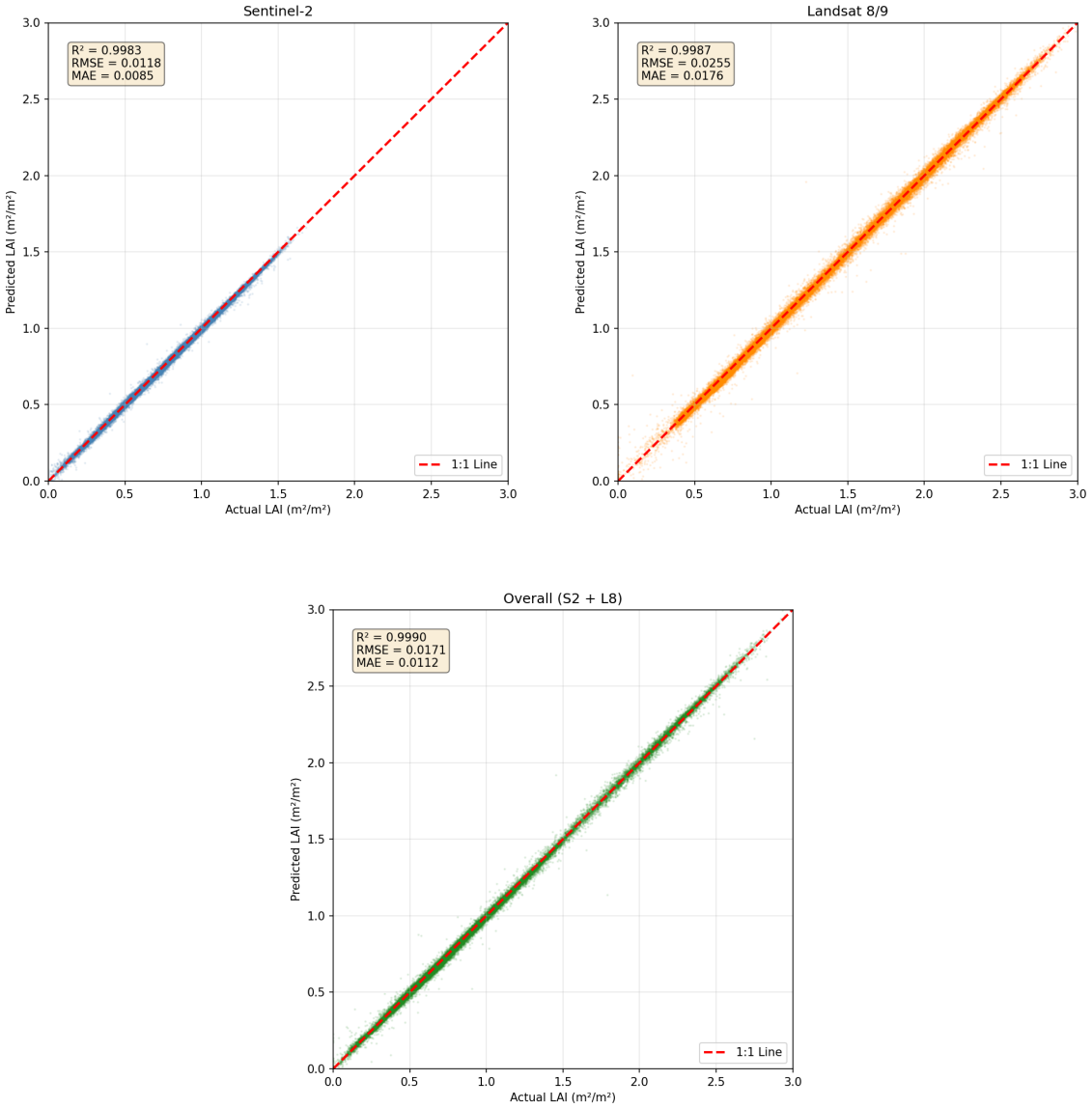

FIGURE 4.3: Scatter plots of predicted versus pseudo-ground truth LAI for Experiment X3.

The model achieves overall $R^2 = 0.999$ and RMSE $= 0.017$ m^2/m^2, with both sensors independently exceeding $R^2 > 0.998$. The model converged in 45 epochs (35 minutes on a single A100 GPU), with training and validation loss curves closely tracking each other without divergence.

Experiment X4: Hybrid VI+PROSAIL Targets

Experiment X4 replaces pure VI-ensemble ground truth with a 50:50 blend of VI-ensemble and PROSAIL-inverted LAI. Hybrid targets degrade overall RMSE by 35% (Table 4.7). Sentinel-2 performance remains comparable (+2%), but Landsat suffers a 33% RMSE increase due to sensor-specific target mismatch: Sentinel-2 tiles receive hybrid labels while Landsat tiles receive only VI-ensemble labels. This

confirms that **ground truth consistency across sensors** is more important than ground truth sophistication.

TABLE 4.7: 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**

Multi-Sensor Synergy

Table 4.8 quantifies the improvement from multi-sensor integration. The multi-sensor model reduces Sentinel-2 RMSE by 80% (0.059 $\rightarrow$ 0.012) and Landsat RMSE by 64% (0.071 $\rightarrow$ 0.026). This demonstrates genuine multi-sensor synergy: each sensor's predictions improve beyond what that sensor achieves alone. Multi-sensor training forces the shared encoder to learn sensor-invariant LAI features; Sentinel-2 provides fine-scale spatial detail and red-edge sensitivity while Landsat contributes robust broadband signatures. Conditional Batch Normalisation enables sensor-specific feature normalisation while preserving shared semantic representations.

TABLE 4.8: 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%

Band Importance Analysis

Permutation feature importance (Figure 4.4) reveals sensor-adaptive spectral strategies that emerge automatically through multi-sensor training. Sentinel-2 relies primarily on NIR (39.4%) and Red (23.4%), consistent with the biophysical basis of vegetation indices, with the three red-edge bands collectively contributing 8.9%. Landsat, lacking red-edge bands, adopts a different strategy: Red (25.5%) and NIR (24.9%) contribute nearly equally, with substantially more importance allocated to SWIR bands (28.2% combined vs. 7.6% for Sentinel-2) and Green (18.2%

vs. 11.5%), compensating for missing red-edge information through sensitivity to canopy water content and leaf structure.

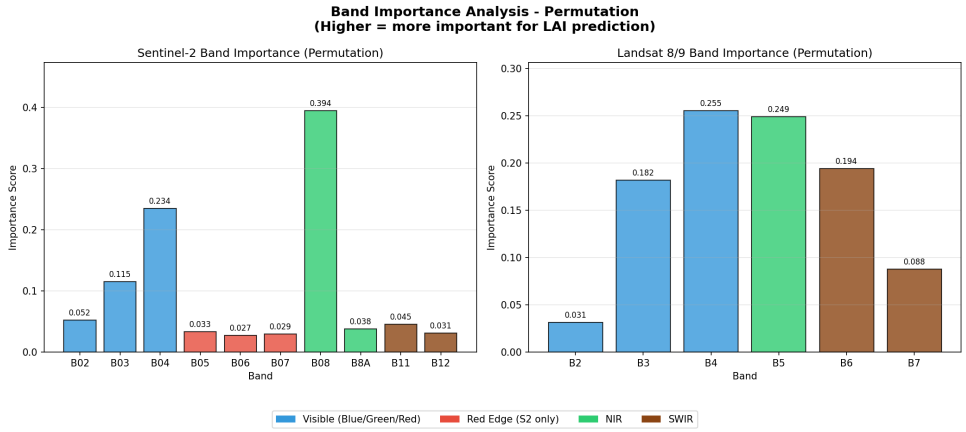

FIGURE 4.4: Permutation-based band importance for Experiment X3.

Summary

Multi-sensor integration provides 80% and 64% RMSE reductions for Sentinel-2 and Landsat respectively. Sensor-agnostic ground truth (pure VI-ensemble) outperforms hybrid VI+PROSAIL targets by maintaining inter-sensor consistency. The model learns sensor-adaptive spectral strategies through the shared encoder-decoder with Conditional Batch Normalisation. These results, validated on nearly 20 million geographically disjoint test pixels, establish the model's capacity for accurate LAI retrieval within the training domain.

4.3 Generalisation Validation

Strong in-domain performance under spatial block validation (Section 4.2) is necessary but insufficient for operational confidence. This section evaluates the best multi-sensor model (X3) across three independent generalisation axes: geographic transfer to an independent region 200 km away, temporal transfer across an 8-month seasonal gap, and methodological comparison against the individual vegetation index components.

Cross-Region Geographic Generalisation

The Burgas region (42.5°N, 27.5°E), approximately 200 km east of Plovdiv, serves as a fully independent validation site with distinct agricultural composition, higher soil salinity and maritime climatic influence. The model was applied without any retraining or calibration – a true zero-shot geographic transfer. Both sensors achieve identical RMSE $= 0.051$ m^2/m^2 (Figure 4.5), well within the VI-ensemble uncertainty bounds ($\sim$0.34 m^2/m^2). This sensor parity confirms that multi-sensor training produces equally transferable representations for both instruments despite their different spectral configurations and spatial resolutions.

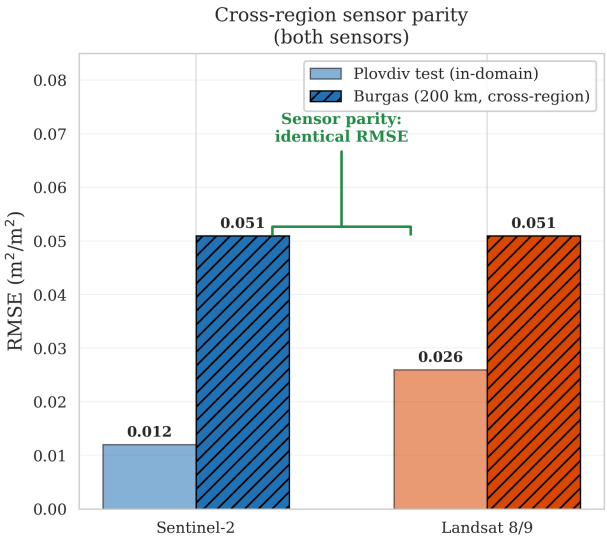

FIGURE 4.5: Cross-region generalisation results demonstrating geographic transferability and sensor-agnostic generalisation of Experiment X3.

Cross-Season Temporal Generalisation

The Burgas region was imaged in both autumn (September 2023) and spring (May 2024), enabling a controlled comparison that isolates temporal effects. Both sensors consistently predict higher LAI in spring (Sentinel-2: 2.9×, Landsat: 1.7×; Figure 4.7), correctly capturing the expected phenological pattern: spring acquisitions correspond to active vegetative growth (LAI 0.6-1.7 m^2/m^2) while autumn captures post-harvest stubble and senescent vegetation (LAI 0.2-1.0 m^2/m^2). The difference in seasonal ratio reflects Sentinel-2's greater sensitivity to fine-scale canopy structure at 10 m resolution with red-edge bands. No systematic bias or drift is observed, confirming that the learned spectral-LAI mapping captures fundamental biophysical relationships rather than season-specific patterns.

CNN vs. Vegetation Index Baseline

Since the CNN is trained on VI-ensemble pseudo-labels, a key question is whether it adds value beyond the constituent vegetation indices. Figure 4.6 compares each method against the VI-ensemble reference on the Plovdiv Sentinel-2 scene (T34TGM). The CNN achieves RMSE $= 0.010$ m^2/m^2, a **77% improvement** over the best individual VI (GNDVI-exp, 0.044). Individual VIs exhibit large inter-method disagreement (RMSE 0.044-0.088) and substantial bias (± 0.05), while the CNN reproduces the ensemble consensus with near-zero bias. This superiority arises from spatial context integration (the U-Net processes 158×158 patches rather than individual pixels), learned non-linear spectral combinations across all 10 bands, noise suppression through learned spatial-spectral regularisation, and

implicit cross-sensor consistency enforced by the shared encoder-decoder.

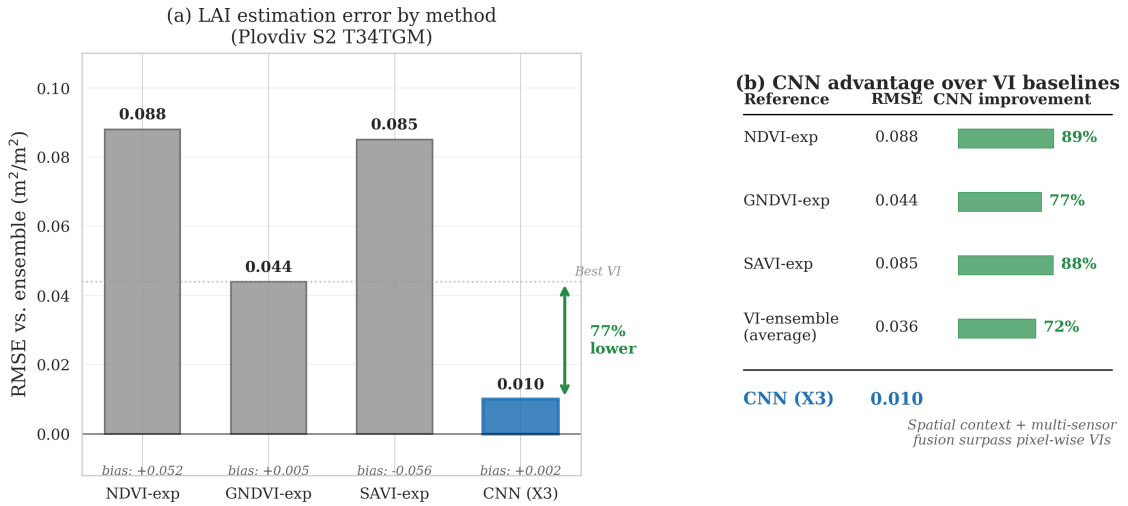

FIGURE 4.6: CNN vs. vegetation index comparison.

Generalisation Summary

Figure 4.7 consolidates performance across all evaluation dimensions. The three axes together establish a triple generalisation achievement: spatial block validation confirms the model does not memorise local patterns, 200 km cross-region transfer confirms geographic robustness and 8-month cross-season transfer confirms temporal stability.

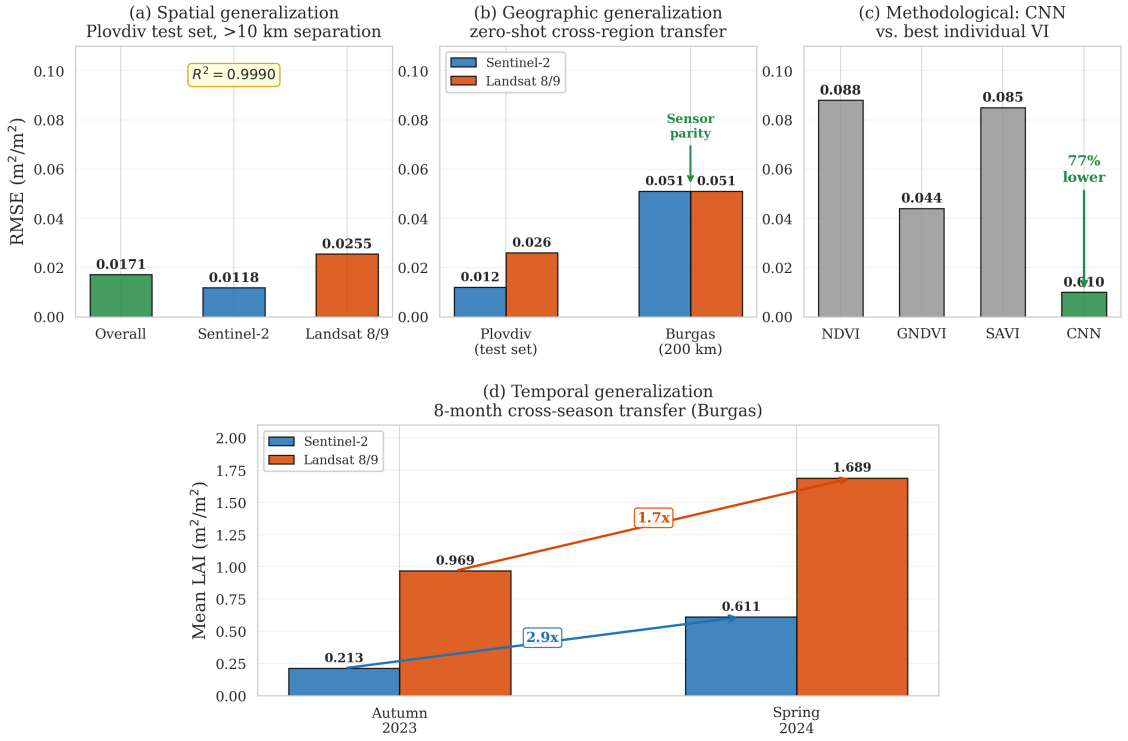

FIGURE 4.7: Triple generalisation framework summary.

Several caveats accompany these results. The VI-ensemble ground truth inherits saturation at high LAI (>5 m^2/m^2) and sensitivity to soil background effects. Geographic generalisation has been validated within temperate agricultural regions of southeastern Europe; performance in tropical, arid or forested landscapes remains untested. Temporal validation spans a single 8-month gap; multi-year stability requires longer time series evaluation.

4.4 Operational Validation

Operational deployment demands full-scene inference at native resolution and interpretable comparison to established LAI products. This section validates the CNN at scene scale against its VI-ensemble training targets and compares predictions to the ESA SNAP Biophysical Processor as an independent reference.

Full-Scene CNN vs. VI-Ensemble

The trained model (X3) was applied to all Plovdiv scenes at full resolution: Sentinel-2 tiles T34TGM and T35TLG (3,374 tiles) and Landsat 9 scene LC09_183031 (14,145 tiles), totalling 17,519 tiles. The CNN achieves near-perfect reproduction of its training targets (Table 4.9), with $R^2 > 0.999$ for both sensors, negligible bias (<0.002 m^2/m^2) and variance ratio ≈1.0, confirming that the full LAI dynamic range is preserved without compression. Full-scene RMSE (0.004–0.010 m^2/m^2) is well below the inter-VI disagreement (∼0.036 m^2/m^2).

TABLE 4.9: 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**

Visual inspection (Figures 4.8 and 4.9) reveals improved spatial coherence in CNN predictions compared to pixel-wise VI-ensemble maps, which exhibit salt-and-pepper noise in heterogeneous areas. The U-Net's spatial receptive field acts as a learned filter, suppressing incoherent pixel-level fluctuations while preserving genuine field-scale variability and sharp boundary transitions.

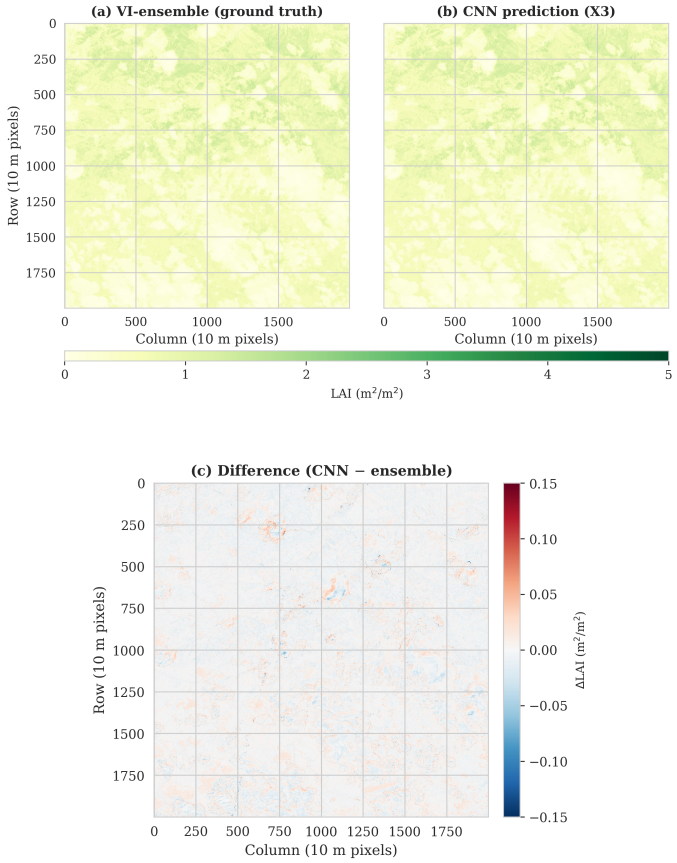

FIGURE 4.8: Representative 20×20 km subregion of Plovdiv tile T34TGM. LAI maps comparing VI-ensemble (left), CNN prediction (centre) and pixel-wise difference (right).

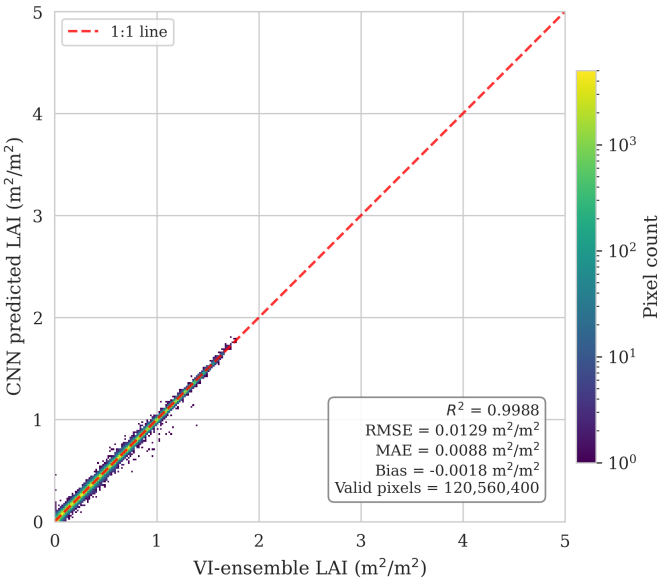

FIGURE 4.9: Density scatter plot of CNN predictions versus VI-ensemble targets across Plovdiv tile T34TGM.

Comparison to ESA SNAP Biophysical Processor

The SNAP Biophysical Processor (v2.1, SNAP 9.0) provides a globally-calibrated LAI product based on an ANN trained on PROSAIL simulations. Figure 4.10 compares the three estimation approaches over the study area. The VI-ensemble produced the most conservative estimates (mean LAI $= 0.55$ m^2/m^2, CV $= 0.49$), SNAP yielded higher values ($\bar{x} = 1.10$ m^2/m^2) with greater spatial variance, and PROSAIL RTM inversion exhibited the widest dynamic range but suffered from inversion instability. Despite differences in absolute magnitude, all methods agreed in delineating vegetated from non-vegetated areas, with the strongest correlation between Ensemble and SNAP ($r_s = 0.87$, $r_p = 0.77$, $p < 0.001$).

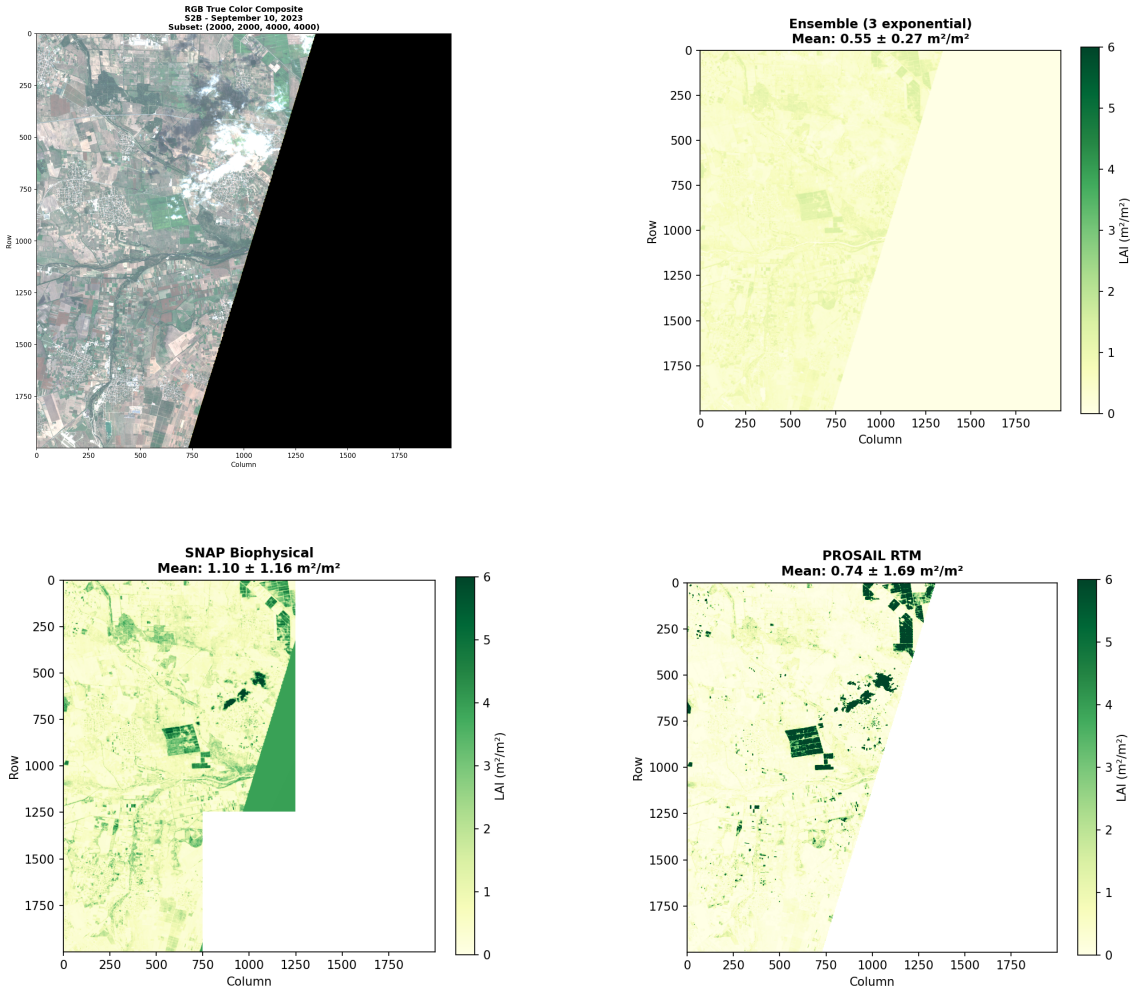

FIGURE 4.10: Spatial comparison of LAI estimation methods for Sentinel-2B scene T35TLG.

With respect to absolute values, however, the two methods diverge substantially (Table 4.10, Figure 4.11): RMSE $= 1.987$ m^2/m^2, $R^2 = 0.312$, which is expected rather than indicative of failure. SNAP predicts, on average, higher LAI ($\bar{x} = 1.10$

compared to 0.55 m^2/m^2 for VI-ensemble), with the discrepancy being most pronounced in dense vegetation, where the VI-ensemble saturates while SNAP continues to increase. Variance ratio = 0.421 indicates a wider dynamic range of SNAP compared to the VI-ensemble.

TABLE 4.10: 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 divergence reflects different LAI definitions and calibration paradigms, not model failure.

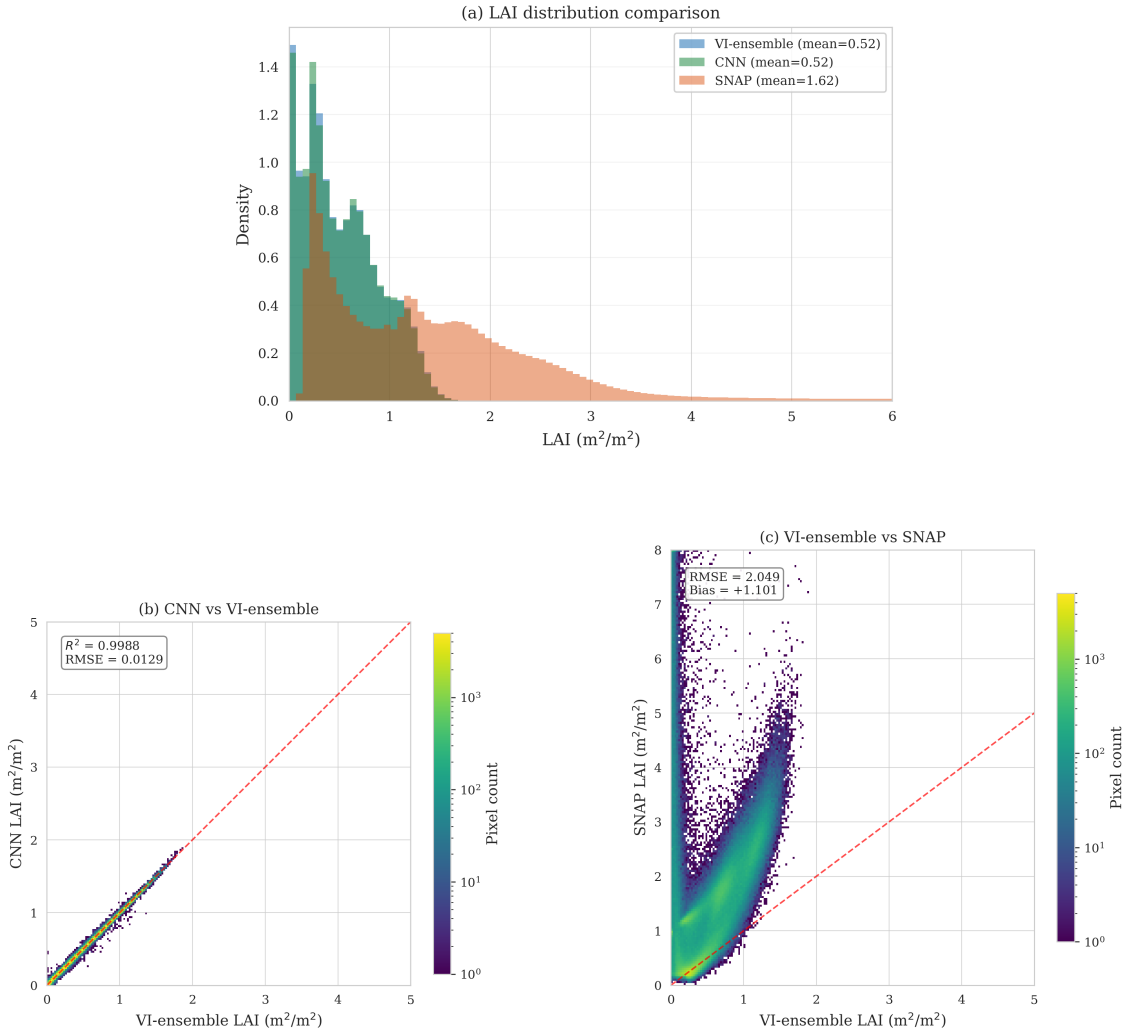

FIGURE 4.11: Operational product comparison.

This divergence reflects four factors: (1) different LAI definitions – SNAP estimates effective LAI (including non-green plant elements), while the VI-ensemble estimates green LAI; (2) different calibration paradigms – SNAP uses globally-calibrated PROSAIL simulations, the VI-ensemble uses empirical agricultural relationships; (3) different saturation mechanisms; and (4) SNAP's pixel-wise processing versus the CNN's patch-based spatial regularisation.

The CNN faithfully reproduces its training paradigm. The two approaches serve complementary purposes: SNAP provides a globally-consistent, physics-grounded product requiring no local calibration, while the CNN offers locally-optimised, spatially-coherent predictions exploiting multi-sensor synergy. Reconciling these paradigms would require field-measured LAI as an independent reference.

4.5 Scalable High-Performance Computational Framework for Full-Scene LAI Mapping

All model training, inference, and large-scale data processing were conducted on the **HEMUS** petascale supercomputer, enabling efficient processing of full-resolution multispectral satellite scenes. This work utilised a single GPU compute node with the specifications in Table 4.11.

TABLE 4.11: Hardware specifications of the HEMUS compute node.

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

The software stack comprised PyTorch 2.1.2 (CUDA 12.1), Python 3.11.7, and supporting libraries (NumPy, Rasterio, GDAL, Scikit-learn), managed through a Conda environment for reproducibility. ESA SNAP was used for biophysical product generation.

Training Efficiency

The A100's Tensor Cores support mixed-precision arithmetic (FP16 into FP32), enabling Automatic Mixed Precision training that doubles throughput without measurable numerical instability. Experiment X3 (28.65M parameters) completed in 35.1 minutes, converging at epoch 29 of 45 (early stopping with patience 15) at approximately 1,500 tiles/second. This rapid cycle enabled the four main experiments and ablation studies to be completed within a single allocation session.

Large-Scale Processing

A key component of the proposed framework is a patch-based inference strategy with overlap-aware reconstruction, which ensures spatial continuity and eliminates boundary artifacts typically observed in tile-based processing approaches.

Full-scene inference for a Sentinel-2 granule (10,980 $\times$ 10,980 pixels, $\sim$120 million valid pixels) completed in under two minutes. Dataset tiling from four satellite scenes involved processing approximately 40 GB of raster data into $\sim$50,000 georeferenced tile pairs. The 503 GB system RAM was critical for these operations, as loading co-registered full-resolution rasters simultaneously exceeds typical workstation capacity.

This design is motivated by the need to process full-resolution multispectral scenes that exceed single-node memory limits, requiring streaming-based tile processing rather than full-image loading.

The framework provides a scalable and physically consistent pipeline for full-scene LAI estimation, enabling direct transition from deep learning inference to operational-scale Earth observation products.

Chapter 5 Conclusions

5.1 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.

5.2 Future Directions

Several extensions would strengthen and broaden this work. Incorporating solar and viewing geometry as auxiliary model inputs – or applying explicit BRDF correction – could improve robustness across illumination conditions without requiring full angular normalisation [4, 40]. A contemporaneous cloud-free WorldView-3 scene (8 VNIR bands at 1.24 m, September 2023) is already available for the study region; fine-tuning Experiment X3 on this data by adding a third sensor identity to the Conditional Batch Normalisation framework would test whether spectral-LAI representations learned at 10-30 m transfer effectively to sub-metre resolution, enabling precision agriculture applications. The most critical next step is retraining with field-measured LAI ground truth to quantify the systematic biases of the VI-ensemble approach and transition the framework from operational pseudo-label estimation to fully calibrated biophysical retrieval. Finally, extending PROSAIL forward modelling to Landsat 8/9 via OLI spectral response convolution [2, 21] would enable sensor-agnostic physics-based training targets.

5.3 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
- 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
- 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.

The third article, as of the end of April 2026, has been cited by 2 publications.

5.4 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.

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 construction of 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.

Bibliography

1. 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).
2. 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).
3. 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).
4. 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).
5. Chen, L.-C., Papandreou, G., Schroff, F. & Adam, H. *Rethinking atrous convolution for semantic image segmentation* **in** *arXiv preprint arXiv:1706.05587* Published at CVPR 2018 (2017).
6. 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).
7. 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).
8. De Vries, H. **andothers**. *Modulating early visual processing by language* **in** *Advances in Neural Information Processing Systems (NeurIPS)* (2017), 6594–6604.
9. 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).
10. Dumoulin, V., Shlens, J. & Kudlur, M. *A learned representation for artistic style* **in** *International Conference on Learning Representations (ICLR)* (2017).
11. 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).
12. 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).
13. 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).
14. 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).
15. 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).

16. 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).
17. Huete, A. R. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* **25,** 295–309. https://doi.org/10.1016/0034-4257(88)90106-X (1988).
18. 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).
19. Iglovikov, V. & Shvets, A. TernausNet: U-Net with VGG11 encoder pre-trained on ImageNet for image segmentation. *arXiv preprint arXiv:1801.05746* (2018).
20. Ioffe, S. & Szegedy, C. *Batch normalization: Accelerating deep network training by reducing internal covariate shift* **in** *International Conference on Machine Learning (ICML)* (2015), 448–456.
21. 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).
22. 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).
23. 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).
24. 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).
25. Liu, Y. **andothers**. The fusion of vegetation indices increases the accuracy of cotton leaf area prediction. *Frontiers in Plant Science* **15,** 1357193 (2024).
26. Loshchilov, I. & Hutter, F. *SGDR: Stochastic gradient descent with warm restarts* **in** *International Conference on Learning Representations (ICLR)* arXiv:1608.03983 (2017).
27. Loshchilov, I. & Hutter, F. *Decoupled weight decay regularization* **in** *International Conference on Learning Representations (ICLR)* arXiv:1711.05101 (2019).
28. Louis, J. **andothers**. *Sentinel-2 Sen2Cor: L2A processor for users* **in** *Proceedings of the Living Planet Symposium* (ESA Special Publication SP-740, 2016), 1–8.
29. 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).
30. 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).
31. NASA. *Landsat 8* https://science.nasa.gov/mission/landsat-8/. Accessed 11.01.2026. 2013.
32. NASA. *Landsat 9* https://science.nasa.gov/mission/landsat-9/. Accessed 11.01.2026. 2021.
33. 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).

34. 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).
35. Ronneberger, O., Fischer, P. & Brox, T. *U-Net: Convolutional networks for biomedical image segmentation* **in** *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (Springer, 2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
36. Rustowicz, R. **andothers**. *Semantic segmentation of crop type in Africa: A novel dataset and analysis of deep learning methods* **in** *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (2019), 75–82.
37. 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).
38. 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).
39. 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).
40. 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).
41. 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).
42. 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).
43. 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).
44. 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).
45. 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).
46. Zhang, Z., Liu, Q. & Wang, Y. *Road extraction by deep residual U-Net* **in** *IEEE Geoscience and Remote Sensing Letters* **15** (2018), 749–753. https://doi.org/10.1109/LGRS.2018.2802944.
47. 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).

Contents