

UNIVERSITÀ
DI PAVIA

UNIVERSITY OF PAVIA

FACULTY OF ENGINEERING

DEPARTMENT OF ELECTRICAL, COMPUTER AND BIOMEDICAL ENGINEERING

MASTER'S DEGREE IN COMPUTER ENGINEERING

Deep Learning-Based Land Cover Mapping in the Philippines Using Sentinel-1 SAR Data and a MOLCA-Derived Training Dataset

Mappatura della copertura del suolo nelle Filippine basata sul deep learning mediante dati SAR Sentinel-1 e un dataset di addestramento derivato da MOLCA

Candidate: **Seyed Amin Fattahzadeh**

Supervisor: **Prof. Paolo Gamba**

Co-supervisors: **Ing. Luigi Russo & Ing. Tonia Sorriso**

A.Y. 2025/2026

Abstract

Reliable high-resolution land-cover mapping using spaceborne Earth Observation data in the Philippines is challenging because persistent cloud cover limits consistent optical observations, while existing land-cover products differ in spatial resolution, legend design, and thematic accuracy. This suggests the use of radar systems, and this thesis develops a complete radar-based workflow using the European Sentinel-1 satellite constellation for producing and evaluating land-cover maps by training a deep learning architecture on a harmonised multi-source reference dataset.

The reference was obtained from eight global, regional, and national land-cover products, harmonised through a Philippines-adapted Map of Land Cover Agreement (MOLCA) procedure. To cover the country, sixty tiles were processed, from which 280 reference patches were selected to ensure broad geographic coverage and rare-class representation. For each scene, 28 Sentinel-1 features derived from four seasonal intervals were retained. The implemented deep learning architecture was a SwinUNETR model trained using a fixed tile-based split and class-weighted cross-entropy.

The final clean baseline achieved an Overall Accuracy of 0.8333, a macro-F1 of 0.5460, a mean Intersection over Union of 0.4339, and a Cohen's kappa of 0.7703 on the fixed internal validation set. "Forest", "Water", and "Cropland" were the strongest classes, whereas "Shrubland" remained unresolved. A source-level audit showed that the "Shrubland" class, which applies only to a small fraction of the country area, had been suppressed during reference construction. Even when a weighted random sampler was used to increase the number of its samples, the improvement was marginal.

The results show that reliable land-cover mapping depends not only on model design and class balancing, but also on reference-data construction, class separability, and geographic consistency.

Riassunto

La produzione di mappe della copertura del suolo ad alta risoluzione nelle Filippine mediante dati satellitari di Osservazione della Terra è particolarmente complessa, poiché la persistente copertura nuvolosa limita la disponibilità di osservazioni ottiche consistenti, mentre i prodotti di copertura del suolo esistenti differiscono per risoluzione spaziale, struttura della legenda e accuratezza tematica. Ciò suggerisce l'impiego di sistemi radar, e questa tesi sviluppa un flusso di lavoro completo basato sui dati radar della costellazione europea Sentinel-1 per la produzione e la valutazione di mappe di copertura del suolo, mediante l'addestramento di un'architettura di deep learning su un dataset di riferimento multi-sorgente armonizzato.

Il riferimento è stato ottenuto da otto prodotti di copertura del suolo globali, regionali e nazionali, armonizzati mediante una procedura Map of Land Cover Agreement (MOLCA) adattata alle Filippine. Per coprire il Paese, sono stati elaborati sessanta tile, dai quali sono state selezionate 280 patch di riferimento per garantire un'ampia copertura geografica e una migliore rappresentazione delle classi rare. Per ogni scena sono state mantenute 28 feature Sentinel-1 derivate da quattro intervalli stagionali. L'architettura di deep learning implementata è un modello SwinUNETR, addestrato utilizzando una suddivisione fissa basata sui tile e una cross-entropy pesata per classe.

La baseline finale ha raggiunto un'Overall Accuracy di 0,8333, un macro-F1 di 0,5460, un mean Intersection over Union di 0,4339 e un coefficiente kappa di Cohen di 0,7703 sul set interno di validazione fissato. Le classi "Forest", "Water" e "Cropland" hanno ottenuto i risultati migliori, mentre "Shrubland" è rimasta irrisolta. Un'analisi a livello delle sorgenti ha mostrato che "la classe 'Shrubland', che interessa solo una piccola frazione del territorio nazionale", era stata fortemente ridotta durante la costruzione del riferimento. Anche utilizzando un campionatore casuale pesato per aumentare il numero dei suoi campioni, il miglioramento è risultato marginale. I risultati mostrano che una mappatura affidabile della copertura del suolo dipende non solo dalla progettazione del modello e dal bilanciamento delle classi, ma anche dalla costruzione dei dati di riferimento, dalla separabilità tra le classi e dalla coerenza geografica.

Acknowledgements

This thesis marks the end of a demanding and deeply formative period of my life. What began as a technical research project gradually became something more personal: an exercise in patience, persistence, responsibility, and learning how to continue when the next step was not obvious. For that reason, I would like to acknowledge not only the people who contributed to the scientific work, but also those who made it possible for me to reach its end.

My first and deepest academic thanks go to my supervisor, **Prof. Paolo Gamba**. I am grateful for the trust he placed in me, for giving me the opportunity to work on a real and challenging remote-sensing problem, and for allowing me the independence to learn through the project rather than simply follow a predefined path. His scientific direction helped me understand that a strong result is not only a good number at the end of an experiment, but a methodology that can be questioned, reproduced, defended, and improved.

I am grateful to my co-supervisors, **Luigi Russo** and **Tonia Sorriso**. Their contribution went far beyond occasional technical advice. Luigi's guidance was fundamental in understanding the deep-learning pipeline, the logic of the original software, the training strategy, and the importance of preserving reproducibility while adapting an existing research system to a new geographic context. Tonia's support was equally important throughout the construction of the Philippine reference dataset, the spatial processing, validation, patch preparation, and Sentinel-1 workflow. I am thankful to both of them for their patience, their detailed feedback, and for taking my questions seriously even when the problems were difficult or required several attempts to solve.

I also thank the **Telecommunications and Remote Sensing Laboratory of the University of Pavia** for providing the scientific environment, computational resources, and technical foundation that made this work possible. Working within this context taught me how much research depends on careful organisation, traceable decisions, and the willingness to verify what appears correct instead of simply assuming that it is.

I am also appreciative of the **Agenzia Spaziale Italiana (ASI)** for selecting me to participate in the ASI Educational Module at the University of Pavia, focused on Earth Observation and its environmental applications. This experience strengthened my interest in space-based technologies and highlighted the importance of connecting technical research with real environmental and societal needs.

There were many moments during this thesis when the work did not proceed as expected: datasets disagreed, classes disappeared from the reference, processing steps had to be reconsidered, models failed on rare classes, and apparently small technical details became important scientific questions. Looking back, these difficulties became one of the most valuable parts of the experience. They taught me that research is not a straight path from an idea to a successful result. It is the discipline of staying with a problem long enough to understand what the result is actually saying.

On a personal level, I am deeply grateful to my family and to the people who remained close to me throughout this period. Their support cannot be represented in a table, a figure, or an evaluation metric, but it is present behind every page of this thesis. They gave me encouragement when I was tired, perspective when the work became overwhelming, and the freedom to keep moving forward. Some journeys are completed individually, but they are never completed alone.

Finally, I want to acknowledge the version of myself who began this work without knowing how many problems would have to be solved before it could be finished. Completing this thesis required learning to accept uncertainty, to start again when necessary, and to value careful progress over easy answers. I hope that the work presented here reflects not only what I learned about remote sensing and deep learning, but also the way I learned to approach difficult problems: with curiosity, precision, humility, and persistence.

"I sought the light of knowledge in the midst of darkness, for there is no dead end on the path of wisdom."
Avicenna

During the drafting and editing of this thesis, I used ChatGPT by OpenAI, Grammarly, SciSpace, and Undermind to support language revision, LaTeX troubleshooting, and literature discovery. These tools were not used to generate original scientific results or analyses. The research questions, methodological decisions, experiments, figures, interpretation of results, and conclusions presented in this thesis are my own. All tool-assisted suggestions were reviewed and edited by me, and citations and factual statements were checked against the original sources.

Contents

Abstract **2**

Riassunto **3**

Acknowledgements **4**

List of Figures **10**

List of Tables **17**

Acronyms and Notation **20**

1 Introduction and Motivation **23**
1.1 Background 23
1.2 Importance of Land Cover Mapping 26
1.3 Challenges in Tropical Regions such as the Philippines 29
1.4 Sentinel-1 SAR for Land Cover Monitoring 34
1.5 Deep Learning for Semantic Segmentation 37
1.6 Aim and Objectives of the Thesis 41
1.7 Thesis Structure 45

2 Study Area and Datasets **47**
2.1 Study Area: The Philippines 47
2.2 Sentinel-2 Tiling Grid Used as Spatial Reference 49
2.3 Input Products for the MOLCA-Derived Reference 51
2.4 Regional Land-Cover Products 57
2.5 Comparison of the Dataset Roles 61
2.6 Final Target Legend 64
2.7 Sentinel-1 SAR Data 66
2.8 Coordinate Reference Systems and Spatial Resolution 67
2.9 Chapter Summary 69

3 Generation of the MOLCA Philippines Reference Map 70
3.1 Purpose of the Chapter . . . 70
3.2 Terminology: AnA and NaN . . . 71
3.3 Conceptual Basis of MOLCA . . . 71
3.4 From General MOLCA to MOLCA Philippines . . . 72
3.5 Tile-Based Processing Strategy . . . 74
3.6 Input Layers Used in the Reference-Map Generation . . . 75
3.7 Preprocessing Requirements . . . 76
3.8 Common Legend and Reclassification Rules . . . 79
3.9 Mathematical Notation for the Agreement Map . . . 85
3.10 Agreement Hierarchy . . . 86
3.11 Workflow Diagram . . . 89
3.12 Quantitative Evidence from AnA Reduction Reports . . . 90
3.13 Staged AnA Reduction Figure . . . 93
3.14 Visual Inspection of Final MOLCA Tile Outputs . . . 94
3.15 Logic Validation and Quality Assurance . . . 96
3.16 Role of AnA in the Final Reference Map . . . 97
3.17 Automation and Reproducibility . . . 97
3.18 Boundary with the Later Shrubland Rescue Experiment . . . 99
3.19 Outputs of Chapter 3 . . . 99
3.20 Limitations of the Reference-Map Generation . . . 100
3.21 Chapter Summary . . . 101

4 Patch Selection and Sentinel-1 Feature Preparation 102
4.1 Purpose of the Chapter . . . 102
4.2 Patch Candidate Generation . . . 103
4.3 Selection of the 60 Sentinel-2 Tiles . . . 106
4.4 Patch Selection Strategy . . . 107
4.5 Class Distribution and Geographic Distribution . . . 111
4.6 Final 280 Selected Patches . . . 113
4.7 Sentinel-1 Data Collection for 2019 . . . 115
4.8 Seasonal Feature Extraction . . . 117
4.9 Generation of Sentinel-1 Feature Stacks . . . 120
4.10 Google Earth Engine Implementation . . . 122
4.11 Alignment between Reference Patches and Sentinel-1 Features . . . 125
4.12 Chapter 4 Workflow Summary . . . 128
4.13 Limitations and Practical Notes . . . 129
4.14 Chapter Summary . . . 131

5 Deep Learning Classification Methodology 132
5.1 Semantic Segmentation Framework . . . 133
5.2 U-Net / Swin-UNETR Model Architecture . . . 138
5.3 Input Features and Output Classes . . . 144
5.4 Training, Validation, and Testing Strategy . . . 151
5.5 Loss Function and Optimisation . . . 161
5.6 Accuracy Assessment Metrics . . . 169
5.7 Chapter Summary . . . 176
Chapter Summary . . . 176

6 Results and Discussion 178
6.1 MOLCA Philippines Reference Dataset . . . 179
6.2 Selected Patch Dataset . . . 185
6.3 Sentinel-1 Feature Dataset . . . 189
6.4 Model Training Results . . . 190
6.5 Segmentation Maps . . . 196
6.6 Accuracy Assessment . . . 201
6.7 Discussion of Errors and Limitations . . . 212
6.8 Chapter Summary . . . 222
Chapter Summary . . . 222

7 Conclusions 224
7.1 Summary . . . 225
7.2 Main Findings . . . 229
7.3 Limitations . . . 236
7.4 Future Work . . . 239

A Complete MOLCA Harmonisation Tables 246
A.1 Common target legend . . . 247
A.2 Original global and thematic products . . . 247
A.3 NAMRIA 2020 harmonisation . . . 249
A.4 JAXA Southeast Asia harmonisation . . . 250

B MOLCA Processing, Quality Control, and Validation Evidence 252
B.1 Tile-level processing sequence . . . 252
B.2 Stage-wise AnA reduction . . . 253
B.3 Logic-validation workbooks . . . 254

C Patch Selection and Sentinel-1 Feature Preparation Evidence 256
C.1 From selected patches to model-ready scenes . . . 257
C.2 Final GEE-ready patch representation . . . 257

C.3 Sentinel-1 acquisition and feature configuration 258
C.4 Model-ready directory organisation 259

D Deep-Learning Reproducibility Manifest **260**
D.1 Final clean-baseline configuration 261
D.2 Final CC50 class weights 261
D.3 Frozen-file checksums 262
D.4 Experiment-artifact organisation 262

E Supplementary Evaluation Tables **264**
E.1 Clean baseline versus 10× sampler 264
E.2 Per-class clean-baseline results 264
E.3 Representative scene-level results 265
E.4 Shrubland evidence retained for interpretation 266

F Project Evidence and Reproducibility File Manifest **267**

Bibliography **269**

List of Figures

1.1 Conceptual organisation of the thesis workflow, from selection and harmonisation of the reference sources to MOLCA-derived reference generation, patch selection, Sentinel-1 feature preparation, and deep-learning classification. Each processing stage is described in detail in the subsequent chapters. 26
1.2 Geographic context of the Philippine study area: (a) Google Satellite basemap viewed through QGIS QuickMapServices; (b) Sentinel-2 tiling framework used to organise the operational study area and subsequent tile- and patch-based processing. 30
1.3 Conceptual comparison of passive optical and active SAR sensing. Optical sensors depend on solar illumination, whereas Sentinel-1 SAR provides its own microwave illumination and can acquire data day or night under most weather conditions. Illustration created for this thesis. 34
1.4 Simplified semantic-segmentation formulation used in the Philippine experiments. Each spatial sample is represented by 28 Sentinel-1-derived feature channels and processed by the 2-D SwinUNETR model to produce a ten-code pixel-wise output comprising NoData/unclassified (ID 0) and nine land-cover classes (IDs 1–9). 41

2.1 Philippine study area and Sentinel-2 tile coverage intersecting the national extent. The GADM administrative polygons provide geographical context, while the Sentinel-2 grid defines the spatial framework used for tile-based processing. 48
2.2 Sentinel-2 tiling framework used in the Philippine workflow. The 60 operational tiles retained for MOLCA processing and subsequent Sentinel-1 feature preparation are highlighted in cyan, while the additional tiles intersecting the broader Philippine extent are shown in pink. 50
2.3 Master coverage views of the two general-purpose global land-cover products used to construct the base agreement of the Philippine MOLCA-derived reference. 53

2.4 Thematic evidence products over a common example area around Manila. The panels illustrate the different spatial and thematic information contributed by the Built-up, surface-water, and forest evidence sources. Each panel uses source-specific symbology; colours and grayscale values should therefore not be interpreted as equivalent class codes across products. . . 56
2.5 NAMRIA land-cover product prepared as a master VRT for the Philippine workflow. The figure illustrates the national coverage used before tile-level clipping and harmonisation. 59
2.6 JAXA Southeast Asia high-resolution land-use/land-cover product prepared as a master VRT for the Philippine workflow. The product provides regional land-cover information designed for Southeast Asian landscapes and complements the global MOLCA inputs. 60
2.7 Example visual comparison over the Manila area for selected general and national/regional products. The purpose of the figure is not to validate the datasets visually, but to illustrate differences in spatial detail, thematic representation, and class structure over the same landscape. Each panel uses source-specific symbology, so colours should not be compared directly between products. 63

3.1 Example tile-grid evidence used during MOLCA processing. The figure documents the 10980 × 10980 pixel structure of a Sentinel-2 tile at 10 m resolution and supports the tile-based processing design used in the thesis. 75
3.2 Functional grouping of the eight reference sources. The diagram shows the role of each dataset family and does not represent the sequential priority of the agreement hierarchy, which is shown in Figure 3.5. 76
3.3 Examples of clipped, 10 m aligned, and harmonised/reclassified layers over the four-tile test area. The figure compares the two general global products with the national NAMRIA product and regional JAXA Southeast Asia product used in the Philippine adaptation. 78
3.4 Examples of clipped, 10 m aligned, and harmonised/reclassified thematic evidence layers. GHS-BUILT-S and WSF support Built-up, GSW supports the project Water/Wetland mapping, and FNF supports Forest decisions. 79
3.5 Complete MOLCA Philippines reference-map generation workflow. The diagram separates raw datasets, spatial preprocessing, semantic harmonisation, agreement computation, national/regional adaptation, quality checking, and final output production. 90
3.6 AnA percentage after each MOLCA processing stage for four example tiles. The staged pattern shows how base agreement, thematic evidence, Mangrove extraction, and final regional agreement progressively reduce unresolved pixels. 93

3.7 Example standard final MOLCA tile outputs before the later Shrubland rescue experiment. These quicklook maps complement the numerical AnA reduction reports by showing spatial patterns in the final agreement outputs. 95

4.1 Schematic 20 × 20 candidate-patch grid for Sentinel-2 tile 51PTR. The highlighted cells correspond to the three patches selected from this tile: R09C15, R10C15, and R10C16. Each grid cell corresponds to an exact 549 × 549 pixel MOLCA reference patch. 105
4.2 The three selected MOLCA-derived ground-reference patches from Sentinel-2 tile 51PTR. The arrangement corresponds to the grid positions highlighted in Figure 4.1. The figure shows the actual categorical reference information subsequently paired with the Sentinel-1 feature rasters. . 106
4.3 National-scale overview of the final 280 selected MOLCA-derived ground-reference patches across the 60 operational Sentinel-2 tiles. At this scale, the figure primarily illustrates the geographical coverage of the selected training areas across the Philippine study area. 112
4.4 Progressive spatial inspection of the final selected MOLCA-derived patch dataset. The two panels are enlarged views of the same 280-patch selection and illustrate how the national-scale sampling pattern translates into individual land-cover reference patches. 112
4.5 Examples of three Sentinel-1 features derived from the first seasonal Super Image for the selected 51PTR patches: Lee-filtered SI, local maximum, and local minimum. 119
4.6 Additional Sentinel-1 spatial features derived from the first seasonal Super Image: local range (MAXMIN), local mean, and local median. Together with SI, LEE, MAX, and MIN, these form the seven features generated for each season. 119
4.7 Seasonal Sentinel-1 VH Super Images for the three selected patches of tile 51PTR. The four panels correspond to the four temporal groups used in the feature-extraction workflow. Although the spatial structures remain geographically consistent, the SAR response varies across the year, providing temporal information that complements the spatial features extracted from each seasonal Super Image. 121
4.8 Visual verification of the correspondence between the MOLCA-derived ground-reference patch and its Sentinel-1 input raster for patch 51PTR_R09_C15 after local post-processing. The ground-reference layer is rendered using categorical pseudocolour, while the Sentinel-1 raster is displayed in grayscale because it contains continuous backscatter-derived values. Both panels represent the same geographical footprint. 126

4.9 QGIS raster-information panels used to verify the result of local post-processing for patch 51PTR_R09_C15. Both the categorical ground-reference raster and the continuous Sentinel-1 SI raster have dimensions of 549×549 pixels and virtually identical projected extents. The very small numerical differences in the displayed extent coordinates are attributable to floating-point representation and are negligible relative to the 10 m spatial resolution. 127
4.10 Final dataset organisation after local post-processing. The folder structure separates Sentinel-1 input features, MOLCA ground-reference patches, and JSON label metadata. 128
4.11 Complete workflow from MOLCA reference tiles to the 280-patch Sentinel-1 extraction and the final 277 aligned, model-ready scene–reference associations. 129

5.1 Overall deep learning workflow used for Sentinel-1 land-cover segmentation in the Philippines, from the MOLCA-derived reference scenes and 28-channel Sentinel-1 feature stacks to model training, validation, full-scene reconstruction, and accuracy assessment. 138
5.2 Simplified architecture of the 2-D MONAI SwinUNETR implementation used in this thesis. 143
5.3 Representative 28-channel Sentinel-1 input stack alongside the MOLCA-derived ground truth for scene 51PTR_09_15. Seven selected channels from different temporal groups are shown to illustrate the varying SAR representations. The complete network input consists of 28 aligned channels, whose exact sequence is defined by the scene's associated JSON file. 146
5.4 Conceptual structure of the input tensor. The 28 Sentinel-1 channels are derived from seven feature representations calculated across four temporal groups, which are subsequently stacked into a single $28 \times 512 \times 512$ input tensor prior to sub-patching. 148
5.5 Patching strategy used to transform each 512×512 scene into nine overlapping 256×256 network inputs. The 50% overlap (stride of 128 pixels) ensures that the network receives multiple contextual views of the central regions of each scene. 150
5.6 Mean land-cover class proportions for the four K-means clusters used to stratify the scene split. By grouping scenes according to their ten-dimensional class-composition vectors, the split ensures that different dominant landscape types (such as predominantly Forest, Water, or Cropland scenes) are proportionally represented in both the training and validation partitions, avoiding the biases of arbitrary random sampling. . . . 152

5.7 Geographic distribution of the selected Sentinel-1 scenes across the Philippines. The dataset is explicitly partitioned into 249 training scenes (light blue squares) and 28 validation scenes (red hashed squares) utilising the `tile_split.json` configuration. This mapping visualises the geographic outcome of the class-composition stratified split, demonstrating that both data subsets successfully maintain a representative and intermixed spatial distribution across the archipelagic extent. 153
5.8 Effect of the WeightedRandomSampler on training-patch exposure during the Shrubland-focused experiment. By assigning a sampling weight of 10 to any patch containing at least one Shrubland pixel and sampling with replacement, the expected probability of exposing the model to Shrubland during a training step increases from 13.20% to 60.35%. This isolates patch-level exposure without synthetically generating new reference pixels. 158
5.9 Illustration of the valid-pixel masking operation prior to loss calculation. Pixels corresponding to the NoData/unclassified category (class ID 0) are excluded from the target mask, ensuring that unlabelled regions do not contribute to the weighted cross-entropy optimisation. 162
5.10 Comparison of the mean training class composition (Panel A) and the resulting inverse-square-root loss weights (Panel B). The visualisation demonstrates how dominant classes like Forest receive minimal weight, while extremely rare classes like Shrubland and Mangrove Forest are heavily penalised for misclassification to counter the severe pixel-level imbalance. 165
5.11 Training control mechanisms used during the deep learning experiments. Panel A illustrates the cosine learning-rate schedule, decaying from an initial rate of 10^{-4} to a minimum of 10^{-6} over a theoretical maximum of 100 epochs. Panel B outlines the validation-based early-stopping logic, which monitors the validation mIoU and terminates training if no improvement greater than 0.001 is observed for 25 consecutive epochs to prevent overfitting. 167
5.12 Full-scene reconstruction procedure used for final validation. The nine overlapping 256×256 probability maps are remapped to their original spatial coordinates. Overlapping regions are combined via arithmetic averaging of the predicted probabilities before the final argmax decision produces a single 512×512 classification map. This guarantees that every valid reference pixel is counted exactly once during global accuracy assessment. 174

6.1 Quantitative effect of connected-component filtering on the conservative Shrubland rescue. The dual-axis plot illustrates the absolute number of retained pixels (bars) and the retention percentage relative to the raw unfiltered rescue (line) across progressively stricter thresholds. This confirms the steep initial loss of fragmented pixels and supports the selection of CC50 as a balanced intermediate threshold. 183

6.2 Visual comparison of the original MOLCA reference and the CC10, CC50, and CC100 connected-component filtering variants for three representative scenes (50PRC, 51QTA, and 51PWM). The comparison illustrates the progressive removal of small isolated Shrubland components as the minimum connected-component threshold increases. 184
6.3 Provenance of Shrubland pixels in the revised CC50 training reference. Only 14.84% of training Shrubland pixels originated from the original reference, whereas 85.16% were introduced by the conservative rescue procedure. This illustrates the strong dependence of the final Shrubland training set on the revised reference generation strategy. 189
6.4 Chronology of the main experiments and diagnostic steps leading to the final CC50 analysis. The workflow began with the original baseline, followed by a reference audit and the selection of the CC50 connected-component threshold. A clean CC50 baseline was then established, after which a 10× WeightedRandomSampler experiment was tested. Finally, a detailed Shrubland diagnostic was conducted to investigate the remaining failure cases. 193
6.5 Comparison of global performance metrics between the CC50 clean baseline and the 10× Shrubland-patch sampler experiment. The bar chart visually confirms that artificially increasing Shrubland training exposure degrades Overall Accuracy, macro F1, and mean IoU globally, indicating that patch-level exposure alone does not resolve the rare-class learning difficulty. 196
6.6 Qualitative land-cover predictions from the pre-CC50 developmental baseline for four representative validation scenes. Each example shows the reference ground truth, model prediction, and corresponding error map, where red pixels indicate disagreement between the prediction and reference. These maps are included as developmental qualitative results and should not be interpreted as outputs from the final CC50 clean baseline. . 197
6.7 Per-class precision, recall, and Intersection over Union (IoU) for the final CC50 clean baseline. The chart visualises the strong performance of dominant classes such as Water, Forest, and Cropland, contrasting with the lower accuracy metrics observed in the minority classes and the complete failure to recover Shrubland. 202
6.8 Row-normalized confusion matrix for the final CC50 clean baseline. Rows represent the MOLCA-derived reference classes, and columns represent the SwinUNETR predicted classes. NoData is excluded from the principal accuracy interpretation. The matrix reveals strong diagonal performance for Water, Forest, and Cropland, while highlighting the systematic misclassification of minority classes, particularly the complete absorption of Shrubland into neighbouring vegetation categories. 207

6.9 Shrubland prediction outcomes comparing the CC50 clean baseline and the 10× sampler experiment. The chart highlights that while the sampler successfully forces the model to predict class 5 (over 26,000 pixels), the overwhelming majority of these predictions are false positives, with only 39 true positives recovered. 209
6.10 Side-by-side comparison of the row-normalized confusion matrices for the CC50 clean baseline (left) and the 10× WeightedRandomSampler experiment (right). The visual comparison confirms that heavily sampling Shrubland patches does not create a distinct Shrubland diagonal; instead, it redistributes false predictions across neighbouring vegetation categories. 211
6.11 Analysis of the five validation scenes containing Shrubland, combining pixel support, reference provenance, maximum within-patch Shrubland fraction, and the 10× sampler true-positive outcome. The visualisation clearly demonstrates that the only scene producing true positives (51QTU_03_11) relies entirely on original MOLCA labels, whereas the four validation scenes derived exclusively from the CC50 rescue yield zero true positives. 216
6.12 Mean posterior probabilities assigned to true Shrubland pixels by the clean CC50 model and the 10× sampler experiment. The diagnostic confirms that genuine Shrubland generally receives very low class-5 posterior probability, with the Forest class dominating the predictions in most scenes. Furthermore, the 10× sampler fails to systematically increase confidence in the correct class. 218

List of Tables

1.1 Main challenges relevant to high-resolution land-cover mapping in the Philippines and the corresponding methodological responses adopted in this thesis. . . . 33

2.1 Spatial framework adopted for the Philippine land-cover mapping workflow. 51
2.2 Global input products used in the MOLCA-derived Philippine reference workflow. . . . 57
2.3 Regional datasets introduced to adapt the MOLCA-derived reference to the Philippine context. . . . 61
2.4 Conceptual role of each dataset in the Philippine MOLCA-derived reference design. . . . 62
2.5 Final target legend used for the Philippine MOLCA-derived reference map. 65
2.6 Sentinel-1 dataset configuration used as predictor data in the thesis. . . . 67
2.7 Main spatial-resolution considerations for the datasets used in this thesis. 68

3.1 Conceptual adaptation of the MOLCA framework to the Philippine workflow. . . . 74
3.2 Final target legend used during MOLCA Philippines reference-map generation. . . . 80
3.3 Reclassification rules for the six original MOLCA-style input datasets used in the Philippine workflow (Part 1). . . . 81
3.4 Reclassification rules for the six original MOLCA-style input datasets used in the Philippine workflow (Part 2). . . . 82
3.5 NAMRIA 2020 harmonisation used in the MOLCA Philippines workflow. 84
3.6 JAXA Southeast Asia harmonisation used in the MOLCA Philippines workflow. . . . 85
3.7 Staged reduction of unresolved AnA pixels for four example tiles. The project CSV files report these unresolved pixels as `NaN_Pixels` and `NaN_Percentage`. . . . 92
3.8 Overall AnA reduction from base agreement to final MOLCA for the four example tiles. . . . 93
3.9 Logic-validation evidence available from the project spreadsheets. . . . 96

4.1 Tile and candidate-patch structure used in Chapter 4. . . . 107

4.2 Class distribution in the 60-tile MOLCA Philippines reference before final patch selection. . . . 108
4.3 Rare and important class presence across the 60 operational tiles. . . . 109
4.4 Top Shrubland tiles in the 60-tile MOLCA reference. . . . 109
4.5 Patch-count structure of the final 280-patch dataset. . . . 111
4.6 Final fields used for reconstructing patch geometries in Google Earth Engine. 114
4.7 Sentinel-1 configuration used for the Philippine feature-extraction workflow. 116
4.8 Four temporal groups used to divide the Sentinel-1 2019 time series. . . . 117
4.9 Sentinel-1 feature plan used for every selected patch. . . . 120
4.10 Main technical issues encountered during the GEE Sentinel-1 test and how they were handled. . . . 124

5.1 Relationship between the methodology of Russo et al. [9] and the Philippine implementation developed in this thesis. The table describes methodological differences rather than performance differences. . . . 137
5.2 Principal SwinUNETR parameters used in the Philippine experiments. . . . 141
5.3 Conceptual organization of the 28 Sentinel-1 input channels. . . . 145
5.4 Output classes used by the deep learning model. . . . 149
5.5 Scene and patch counts used in the deep learning pipeline. . . . 154
5.6 Principal computational environment used for full model training. . . . 161
5.7 Inverse-square-root class weights used for the final CC50 training experiments. The weights were computed exclusively from the 249 training scenes. . . . 164
5.8 Summary of the principal training configuration used for the final CC50 experiments. . . . 168
5.9 Principal metrics used for evaluation of the Philippine land-cover segmentation model. . . . 176

6.1 Class distribution of the original final MOLCA Philippines reference across the 60 operational Sentinel-2 tiles before the conservative Shrubland revision. . . . 180
6.2 Effect of connected-component filtering on the conservative Shrubland rescue. . . . 183
6.3 Shrubland representation in the final CC50 deep learning dataset. . . . 187
6.4 Provenance of Shrubland supervision in the CC50 training set. . . . 188
6.5 Chronology of the principal deep learning experiments. . . . 192
6.6 Global full-scene unique-pixel validation results for the CC50 clean baseline. 194
6.7 Controlled comparison of the CC50 clean baseline and the 10× Shrubland-patch WeightedRandomSampler experiment. . . . 195
6.8 Selected scene-level results for the final CC50 clean baseline. . . . 201
6.9 Per-class unique-pixel validation results for the final CC50 clean baseline. NoData is excluded from the accuracy assessment. . . . 202
6.10 Reference-class origin of the 26,056 pixels predicted as Shrubland by the 10× sampler model. . . . 210

6.11 Maximum Shrubland support within one 256 × 256 patch for the five Shrubland validation scenes. 214
6.12 Mean clean-model probabilities on true Shrubland pixels. 217

A.1 Final ten-code reference legend used throughout the Philippines workflow. 247
A.2 Complete harmonisation of the six original MOLCA-style input products. 247
A.3 NAMRIA 2020 raster-value mapping used in the final harmonisation. . . . 250
A.4 JAXA 2023SEA v25.09 harmonisation used in the Philippine workflow. . 250

B.1 Archived staged AnA/NaN reduction for four representative Sentinel-2 tiles. 253
B.2 Status of the archived MOLCA logic-validation workbooks. 255

C.1 Main sample counts from spatial selection to deep-learning input. 257
C.2 Fields used to reconstruct selected patch geometries in Google Earth Engine. 258
C.3 Final Sentinel-1 and spatial configuration carried into the model. 258

D.1 Principal frozen configuration of the final clean CC50 baseline. 261
D.2 Class weights frozen for the final CC50 experiments. 261
D.3 Checksums recorded for the frozen clean-baseline reproducibility package. 262

E.1 Controlled comparison of the final clean CC50 baseline and 10× sampler experiment. 264
E.2 Per-class unique-pixel validation results for the final clean CC50 baseline. 265
E.3 Representative full-scene validation results illustrating geographic variation in difficulty. 265

F.1 Principal project evidence retained outside the main thesis narrative. . . . 267

Acronyms and Notation

Acronym	Meaning
AI	Artificial Intelligence
AnA	Area of No Agreement / Area not Available
AOI	Area of Interest
BOM	Byte Order Mark
CC	Connected Component
CE	Cross-Entropy
CNN	Convolutional Neural Network
CRS	Coordinate Reference System
CSV	Comma-Separated Values
CUDA	Compute Unified Device Architecture
DL	Deep Learning
DLR	German Aerospace Center
EO	Earth Observation
ESA	European Space Agency
F1	F1-score
FN	False Negative
FNF	Forest/Non-Forest
FP	False Positive
FROM-GLC	Finer Resolution Observation and Monitoring of Global Land Cover
GADM	Database of Global Administrative Areas
GDAL	Geospatial Data Abstraction Library
GEE	Google Earth Engine
GHS-BUILT-S	Global Human Settlement Built-Up Surface product
GIS	Geographic Information System
GL30	GlobeLand30
GPU	Graphics Processing Unit
GRASS	Geographic Resources Analysis Support System
GRD	Ground Range Detected
GSW	Global Surface Water
GT	Ground Truth / reference labels

Acronym	Meaning
HRLC	High-Resolution Land Cover
IoU	Intersection over Union
IW	Interferometric Wide Swath
JAXA	Japan Aerospace Exploration Agency
JRC	Joint Research Centre
JSON	JavaScript Object Notation
LC	Land Cover
LULC	Land Use / Land Cover
MGRS	Military Grid Reference System
mIoU	mean Intersection over Union
MOLCA	Map of Land Cover Agreement
MONAI	Medical Open Network for AI
NAMRIA	National Mapping and Resource Information Authority
NoData	Invalid or unassigned raster value
OA	Overall Accuracy
PA	Producer's Accuracy
QGIS	Quantum Geographic Information System
S1	Sentinel-1
S2	Sentinel-2
SAR	Synthetic Aperture Radar
SI	Super Image
TP	True Positive
UA	User's Accuracy
UTM	Universal Transverse Mercator
VH	Vertical transmit–Horizontal receive SAR polarization
ViT	Vision Transformer
VRT	Virtual Raster
WCE	Weighted Cross-Entropy
WSF	World Settlement Footprint

Notation

The principal mathematical symbols used in the methodology and evaluation chapters are listed below.

Symbol	Meaning
$\mathbf{X}$	Input tensor supplied to the neural network
$\mathbf{Y}$	Reference-label tensor
$\mathbf{Z}$	Network logits before the final class decision
B	Batch size
C	Number of input channels
H, W	Spatial height and width
K	Number of output classes/codes considered by the model
TP, FP, FN	True-positive, false-positive, and false-negative counts
η	Learning rate
θ	Model parameters
$\mathcal{L}$	Loss function

Chapter 1

Introduction and Motivation

1.1 Background

Land cover describes the observable biophysical material present at the Earth's surface. It includes natural and human-made surface categories such as forests, grasslands, croplands, shrublands, wetlands, water bodies, bare surfaces, and built-up areas. This concept must be distinguished from *land use*. While land cover describes what is physically present on the surface, land use refers to the arrangements, activities, and inputs through which people use, modify, or maintain land [1]. This distinction defines the scope of the present thesis: the objective is to perform *land cover mapping*. The work does not attempt to infer the socioeconomic use of individual land parcels, nor does it formulate the problem as land-cover change detection.

Land cover is one of the fundamental spatial descriptions of the terrestrial environment. A land-cover map transforms remotely sensed observations into a categorical representation in which each mapped location is associated with a meaningful surface class. In mathematical terms, if the geographic domain of interest is represented by a set of spatial locations Ω, a land-cover map can be considered as a function

$$L : \Omega \to \mathcal{C}, \tag{1.1}$$

where $\mathcal{C} = \{c_1, c_2, \ldots, c_K\}$ denotes the set of K land-cover classes. The task of an automated mapping system is therefore to estimate L from remotely sensed observations while preserving both thematic correctness and spatial consistency.

The increasing availability of Earth Observation (EO) satellite data has greatly expanded the possibility of producing such maps over large areas. Modern satellite missions provide repeated observations at spatial resolutions that allow the landscape to be described at substantially finer scales than were possible with many

earlier global land-cover products. Within the terminology adopted by the European Space Agency (ESA) Climate Change Initiative, high-resolution land-cover products are associated with EO observations at approximately 10–30 m spatial resolution [2]. At these scales, considerably more spatial detail can be represented than in traditional products with pixel sizes of several hundred metres.

Higher spatial resolution is particularly important in heterogeneous landscapes. A coarse pixel can contain several surface materials at the same time, for example trees, roads, buildings, cropland, and water. Such a pixel must either be assigned to a dominant class or represented using a more complex mixture model. As the spatial resolution becomes finer, many of these landscape components can be represented more explicitly. This improves the capacity of land-cover products to characterise fragmented forests, small agricultural areas, urban structures, narrow water bodies, coastal features, and transitions between neighbouring surface types. High spatial resolution, however, also increases the amount of data to be processed and raises the demand for detailed and geographically representative reference labels.

This reference-data requirement has become one of the central challenges in large-scale land-cover mapping. Supervised machine-learning and deep-learning algorithms learn a relationship between remotely sensed observations and known class labels. Their performance therefore depends not only on the quality of the satellite imagery and the selected model architecture, but also on the amount, spatial distribution, thematic consistency, and accuracy of the reference data used during training. Pixel-level annotation over large geographic regions is expensive and difficult to perform manually. The problem becomes more pronounced for deep neural networks, which can contain large numbers of trainable parameters and typically benefit from extensive and diverse training samples [3, 4].

One approach to reducing this reference-data bottleneck is to reuse information already contained in existing land-cover products. This idea forms the basis of the Map of Land Cover Agreement (MOLCA) methodology developed by Bratic et al. [3]. Rather than accepting an individual land-cover map as error-free reference information, MOLCA combines multiple existing high-resolution land-cover datasets and retains locations at which the participating products satisfy predefined conditions of thematic agreement. Areas in which the products disagree are excluded from the resulting reference dataset. The original MOLCA was developed for large regions of Africa, Amazonia, and Siberia and provides a very large number of 10 m reference pixels for training land-cover classifiers [3]. The underlying principle is particularly relevant to large-scale deep learning because it converts previously produced land-cover information into a systematic source of candidate reference labels.

The present thesis transfers and adapts this agreement-based philosophy to the

Philippines. The resulting reference dataset is therefore described as *MOLCA-derived*: it follows the central idea of combining and harmonising multiple land-cover products, while introducing modifications required by the geographic context, the available national and regional datasets, and the objectives of the Philippines mapping experiment. The reference-generation process is not treated as a secondary preprocessing step. It constitutes one of the principal methodological components of the thesis because the quality and spatial distribution of the training labels directly influence the subsequent deep-learning classification.

The second principal component is the use of Synthetic Aperture Radar (SAR) Earth Observation. Many high-resolution land-cover mapping workflows rely strongly on multispectral optical imagery because reflectance measured at different wavelengths provides intuitive information about vegetation, water, soil, and artificial surfaces. Optical systems, however, depend on the availability of sufficiently cloud-free observations. Persistent clouds, haze, and atmospheric conditions can create major temporal gaps in humid tropical regions. Radar observations provide an important alternative because microwave sensing does not depend on solar illumination and is largely insensitive to cloud obstruction [5, 6].

The European Copernicus Sentinel-1 mission is particularly suitable for this purpose. Sentinel-1 provides repeated C-band SAR observations and supports systematic monitoring of the Earth's surface [5]. Instead of measuring reflected sunlight, SAR records microwave energy transmitted by the sensor and scattered back from the surface. Radar backscatter therefore contains information related to surface roughness, geometry, dielectric properties, vegetation structure, moisture conditions, wavelength, polarisation, and acquisition geometry [6]. These characteristics make Sentinel-1 fundamentally different from optical imagery and create both opportunities and challenges for land-cover discrimination.

A single SAR acquisition may not contain sufficient information to separate all relevant land-cover classes. Different surface types can produce partially overlapping backscatter responses, while the same class can vary through time because of vegetation development, precipitation, soil moisture, agricultural activity, flooding, or other environmental conditions. Repeated Sentinel-1 observations provide an additional temporal dimension that can help resolve some of this ambiguity. Research in cloud-persistent tropical environments has shown that preserving temporal information in radar image sequences can improve land-cover classification relative to representing the sequence only through a simple temporal average [7]. Recent work has also demonstrated that deep architectures can learn useful spatio-temporal representations directly from Sentinel image sequences [8, 9].

These developments motivate the methodological combination investigated in this thesis: an agreement-derived reference dataset provides the supervision, multitemporal Sentinel-1 SAR observations provide the predictor information, and a

deep semantic-segmentation model converts the SAR feature representation into a spatially continuous land-cover map. Figure 1.1 summarises this relationship at a conceptual level. Each component is developed in detail in the subsequent chapters.

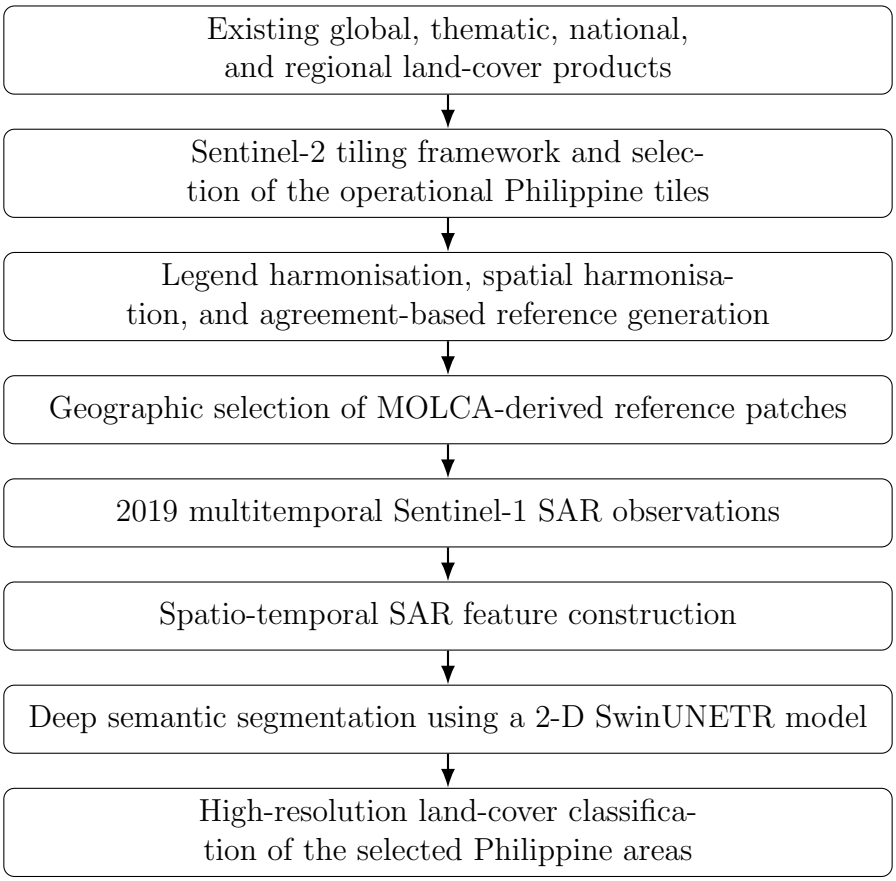


Figure 1.1: Conceptual organisation of the thesis workflow, from selection and harmonisation of the reference sources to MOLCA-derived reference generation, patch selection, Sentinel-1 feature preparation, and deep-learning classification. Each processing stage is described in detail in the subsequent chapters.

1.2 Importance of Land Cover Mapping

Accurate information about the physical state of the Earth's surface is required in many environmental and territorial applications. Land cover controls or influences interactions among vegetation, soil, water, the atmosphere, and human-made surfaces. Differences between forest, cropland, bare soil, wetlands, water, and built-up

areas correspond to different surface properties and therefore to different ecological, hydrological, and biophysical behaviours. For this reason, land-cover information is used in fields ranging from climate and ecosystem analysis to natural-resource management, agriculture, hydrology, urban planning, and disaster-related assessment [3, 2].

The usefulness of a land-cover product depends on both its thematic accuracy and its spatial detail. Thematic accuracy concerns whether a location is assigned to the correct class. Spatial detail concerns whether the map can represent landscape structures at the scale required by an application. These two properties are connected. Increasing spatial resolution does not automatically produce a more accurate map; instead, it exposes finer spatial variability and creates a more demanding classification problem. A high-resolution classifier must distinguish small features and boundaries while maintaining consistent predictions over large homogeneous objects.

The benefit of fine spatial detail is especially evident in landscapes where many classes coexist within short distances. At coarser resolution, a settlement surrounded by cropland may be absorbed into an agricultural class, narrow rivers may disappear, and small forest fragments may be merged with neighbouring vegetation. At approximately 10 m sampling, substantially more of this heterogeneity can be represented. ESA's High Resolution Land Cover initiative uses 10–30 m EO observations precisely because such resolution provides additional information for regional-scale analysis compared with traditional medium-resolution products [2].

For the Philippines, detailed land-cover information has strong practical relevance because terrestrial, coastal, agricultural, forested, and urban environments often occur in close spatial proximity. Local studies in the country demonstrate the importance of carefully selecting and combining EO features when mapping heterogeneous landscapes. For example, Almarines et al. evaluated image compositing and multisource information for multitemporal mapping of two Philippine watersheds and showed that mapping performance depends on the interaction between input features, data fusion, and local landscape characteristics [10]. This reinforces an important point for the present thesis: a method that performs well in one geographic region cannot simply be assumed to behave identically in another.

Land-cover mapping at high resolution also supports the interpretation of spatial patterns that are hidden by more generalised products. A pixel-level map can reveal the configuration and fragmentation of surface classes, the interfaces between vegetation and built-up areas, the spatial continuity of water and wetlands, and the distribution of smaller patches of vegetation. Such information is relevant not only to the final map user but also to the training of machine-learning systems. A reference dataset with insufficient spatial detail may transfer its generalisation

errors to the trained model.

Consequently, map production should not be understood as the final neural network inference step alone. It is an end-to-end process in which source data selection, reference-label construction, spatial alignment, sampling, feature engineering, model design, and evaluation are tightly connected. An error introduced early in this chain can propagate into later stages. For example, a systematic misalignment between a reference raster and a SAR image can create apparently incorrect training samples along boundaries even when both source products are individually accurate. Similarly, two reference datasets may assign different labels to the same physical surface because their class definitions are not equivalent.

Legend harmonisation therefore plays an important role in multi-source high-resolution mapping. Land-cover products produced by different organisations can differ in nomenclature, class definitions, spatial resolution, coordinate reference system, baseline year, and mapping methodology. A label named "forest" in one product is not guaranteed to have exactly the same semantic definition as "forest" in another. The MOLCA methodology addresses this problem by first establishing a common target legend and then identifying only compatible areas of agreement [3]. This idea is adopted and extended in the present study for the Philippines.

The need for reliable reference data becomes still more important when deep learning is used. Deep segmentation networks are capable of learning complex non-linear relationships and spatial context, but the network optimises its parameters according to the labels presented during training. A large dataset of inconsistent labels can therefore teach the network an inconsistent decision function. Conversely, a conservative reference strategy that retains only locations supported by agreement among multiple products can trade spatial completeness for greater agreement-based confidence in the retained labels. The balance between reference quality, spatial coverage, class representation, and geographic diversity is therefore one of the central design considerations of this thesis.

The importance of land-cover mapping in this work can consequently be viewed at three connected levels. At the *environmental level*, the map represents the spatial organisation of physical surface classes. At the *Earth Observation level*, the mapping problem tests the capacity of multitemporal C-band SAR information to discriminate land-cover categories in a cloud-prone tropical country. At the *machine-learning level*, the problem provides a demanding semantic-segmentation task in which reference quality, class imbalance, temporal information, and spatial generalisation must all be considered together.

This combination also gives the research methodological relevance beyond the final classified areas. The workflow investigates whether information already available in existing land-cover products can be systematically transformed into training labels and paired with temporally summarised Sentinel-1 information to support

deep land-cover mapping. The resulting workflow therefore addresses two persistent difficulties simultaneously: the availability of suitable training labels and the availability of consistent satellite observations in areas where optical data are frequently obstructed by clouds.

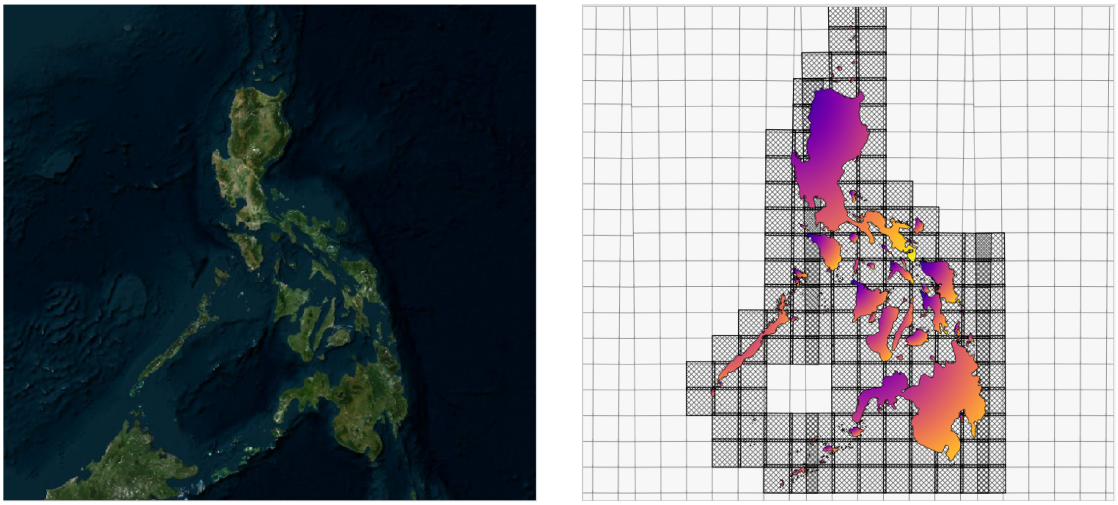
1.3 Challenges in Tropical Regions such as the Philippines

Land-cover mapping in tropical environments introduces a combination of observational and methodological difficulties. The first is the availability of optical satellite observations. Passive optical sensors record reflected solar radiation and therefore require appropriate illumination and a sufficiently unobstructed path between the land surface and the sensor. Clouds and cloud shadows can hide surface information completely, while haze and variable atmospheric conditions can reduce the consistency of usable observations. These effects are particularly important when a mapping workflow requires observations from several dates rather than a single image.

The Philippines presents a strong example of this problem. Its climate is characterised by substantial rainfall and pronounced seasonal and regional differences in precipitation. The Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) distinguishes a rainy season and a dry season at the national level while also recognising several climate types according to the regional distribution of rainfall [11]. Tropical cyclones, monsoon circulation, humidity, and cloudiness further affect the temporal availability and quality of optical imagery. Consequently, assembling a spatially and temporally consistent optical image collection over a large part of the country can require extensive cloud masking, compositing, or gap handling.

The impact of cloud persistence on land-cover classification has been documented in other tropical environments as well. Lopes et al. demonstrated that the temporal dimension of Sentinel-1 and Sentinel-2 observations can provide important information for mapping in a cloud-persistent tropical region, and that maintaining temporal information rather than reducing the entire sequence to simple averages can substantially affect classification performance [7]. Similar motivation applies to the Philippines: repeated microwave observations provide a means of maintaining temporal information even when optical coverage is incomplete.

The second challenge is landscape heterogeneity. Tropical landscapes may contain structurally similar vegetation classes, mixed agricultural systems, wetlands, settlements, open areas, forest edges, and water-related classes within a relatively small geographic area. Some of these classes may have similar instantaneous



(a) Google Satellite basemap of the Philippines

(b) Sentinel-2 tiling framework

Figure 1.2: Geographic context of the Philippine study area: (a) Google Satellite basemap viewed through QGIS QuickMapServices; (b) Sentinel-2 tiling framework used to organise the operational study area and subsequent tile- and patch-based processing.

remote-sensing signatures. Dense vegetation classes, for example, can be difficult to separate using a single acquisition when their structural and moisture characteristics overlap. Agricultural areas can change considerably over the course of a year. Wetland signatures can change according to inundation conditions. Urban and bare surfaces may also produce overlapping responses depending on surface roughness and acquisition geometry.

For radar imagery, this heterogeneity is expressed through the scattering mechanism rather than spectral reflectance. The backscattered signal is affected by surface structure, dielectric properties, geometry, moisture, and vegetation. Two semantically different classes can therefore exhibit similar backscatter values, while the same class can exhibit different values under different environmental or acquisition conditions [6]. This makes single-date, single-feature classification difficult and motivates the use of spatial context and temporal descriptors.

A third challenge concerns terrain. Radar is a side-looking imaging system, so the relationship between the sensor and local topography can strongly influence the measured signal. Terrain changes the effective local incidence angle and can produce geometric effects including foreshortening, layover, and radar shadow in sufficiently steep areas [6]. These effects are different from the cloud-related limitations of optical sensors and illustrate an important principle: SAR does not

eliminate the difficulties of tropical mapping; it replaces some observational limitations with a different set of physical and processing considerations.

A fourth challenge is speckle. SAR is a coherent imaging system, and the interference among echoes originating within a resolution cell gives rise to a granular appearance known as speckle. Although speckle is inherent to the imaging mechanism and contains statistical information, it can obscure spatial patterns and reduce class separability when interpreted directly [6]. Spatial filtering and multitemporal processing can reduce its influence, but excessive filtering can also remove useful spatial details and class boundaries. The treatment of speckle must therefore balance radiometric stability with preservation of spatial information.

The fifth challenge lies in the reference labels. Existing land-cover maps are not interchangeable ground truth. They may have been produced from different satellite sensors, algorithms, legends, acquisition periods, and validation procedures. Their spatial resolutions can also differ. When multiple datasets are combined, a pixel-by-pixel comparison is meaningful only after their coordinate reference systems, pixel grids, extents, resolutions, and class definitions have been harmonised. The original MOLCA study explicitly addresses these issues before computing areas of agreement [3].

Temporal mismatch among reference products creates an additional source of uncertainty. If two maps describe different baseline years, a genuine surface transition that occurred between those years can appear as a disagreement even if both maps correctly represent their respective dates. The MOLCA principle has a useful property in this respect: when disagreement occurs, the affected location can be omitted rather than being automatically accepted as a training label [3]. Nevertheless, agreement does not guarantee that every retained pixel is error-free, and temporal representativeness must still be considered when interpreting the resulting training dataset.

This issue is relevant to the Philippines dataset developed in this thesis, because the contributing products do not all share an identical reference year. The agreement-based procedure is therefore used as a mechanism for constructing a conservative reference dataset rather than as a claim of perfect ground truth. The individual source products, their dates, resolutions, legends, and roles are described in Chapter 2, while the exact harmonisation and agreement rules are presented in Chapter 3.

A sixth challenge is unequal class availability. Land-cover classes are not distributed uniformly across geographic space. A random sample of the Philippines will naturally contain many pixels from widespread classes and far fewer from geographically restricted classes. This imbalance can affect the training of a supervised model, because an optimisation process minimising aggregate classification error can favour common classes unless sampling, weighting, or another imbalance-aware

strategy is introduced. The problem cannot be solved adequately from class counts alone: spatial diversity also matters. A model trained on many neighbouring examples of the same class may generalise less effectively than one trained on fewer but geographically diverse examples.

The seventh challenge is scale. A workflow that is technically successful for a single small study site may become difficult to operate over dozens of satellite tiles and hundreds of patches. Data downloading, processing, file organisation, spatial alignment, feature generation, and task management become part of the scientific workflow. Reproducibility therefore requires that these operational steps be documented alongside the classification model itself.

Table 1.1 summarises the main challenges relevant to the present study and the corresponding methodological responses. The table is intentionally conceptual; the implementation details are introduced in Chapter 2–5.

Table 1.1: Main challenges relevant to high-resolution land-cover mapping in the Philippines and the corresponding methodological responses adopted in this thesis.

Challenge	**Consequence for mapping**	**Response in this thesis**
Persistent cloud cover	Incomplete/inconsistent optical observations	Uses Sentinel-1 SAR for weather-independent, day-and-night observations.
Heterogeneous landscapes	Similar class signatures and complex boundaries	Employs dense semantic segmentation to leverage spatial context.
Temporal surface variability	Single acquisitions miss characteristic behaviors	Organizes repeated S1 data into four temporal groups via spatio-temporal features.
SAR speckle	Reduced interpretability and overlapping responses	Combines multitemporal processing, filtering, spatial statistics, and deep learning.
Inconsistent baseline products	Reference label mismatches (varying legends, years, methods)	Applies a Philippines-specific MOLCA agreement after spatial/thematic harmonisation.
Class imbalance	Minority classes overwhelmed by dominant classes	Controls class distribution during patch selection; details mitigation in chapters.
Large data volume	Manual processing lacks scalability and reproducibility	Adopts scripted preprocessing, tile organisation, and automated patch management.

The important conclusion from these challenges is that there is no single operation that determines the quality of a Philippine land-cover map. Reliable mapping requires a chain of mutually consistent decisions. The choice of Sentinel-1 addresses the observational problem of cloud obstruction, but it introduces the need to interpret radar scattering and manage speckle. Deep learning provides powerful nonlinear and contextual modelling, but it introduces a strong dependence on training-label quality and class representation. Agreement-based reference construction provides a large source of candidate labels, but it requires rigorous spatial and semantic harmonisation. The workflow developed in this thesis is designed around these dependencies rather than treating them as isolated problems.

1.4 Sentinel-1 SAR for Land Cover Monitoring

Synthetic Aperture Radar is an active microwave remote-sensing technique. An active sensor transmits electromagnetic energy toward the Earth's surface and measures a portion of the energy scattered back toward the antenna. This differs fundamentally from passive optical imaging, where the sensor primarily measures solar radiation reflected from the surface. Because SAR provides its own illumination and operates at microwave wavelengths, observations can be acquired during both day and night and under most atmospheric conditions [5, 6].

This conceptual difference is illustrated in Figure 1.3.

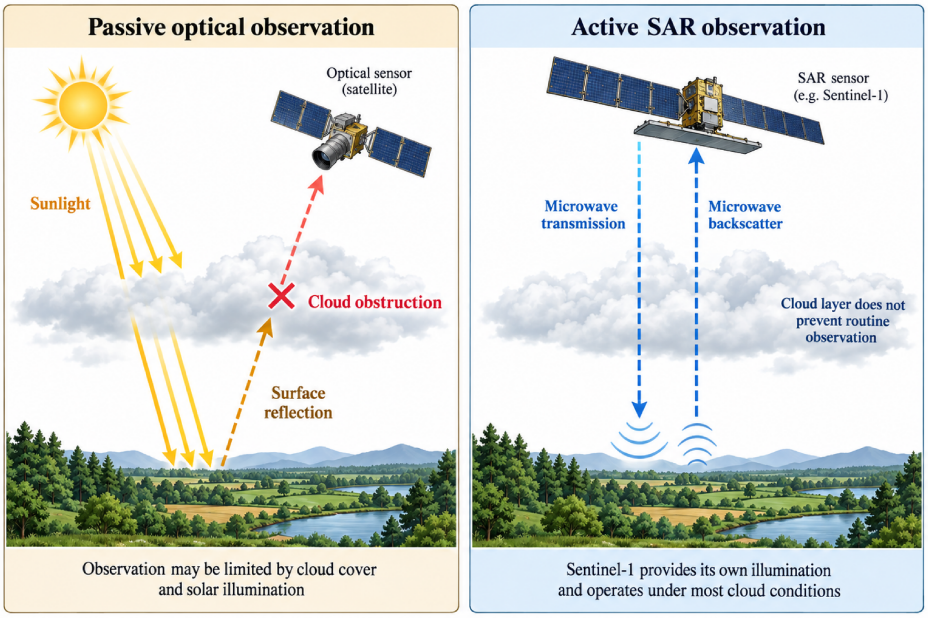


Figure 1.3: Conceptual comparison of passive optical and active SAR sensing. Optical sensors depend on solar illumination, whereas Sentinel-1 SAR provides its own microwave illumination and can acquire data day or night under most weather conditions. Illustration created for this thesis.

The Sentinel-1 mission was developed as a C-band SAR component of the European Earth Observation programme and was designed for systematic, repeated monitoring [5]. Its combination of open data access, repeated acquisition, wide-area imaging, and metre-to-decametre scale sampling has made it an important source for environmental remote sensing. For land-cover mapping, the principal advantage is not that radar directly observes semantic classes, but that different surface structures produce different patterns of microwave scattering that can be learned by a classifier.

A common radiometric quantity used in SAR analysis is the normalised radar backscatter coefficient, σ^0. When represented in decibels, a linear backscatter value can be transformed according to

$$\sigma^0_{\mathrm{dB}} = 10 \log_{10} \left(\sigma^0_{\mathrm{lin}} \right), \tag{1.2}$$

where σ^0_{lin} is the backscatter coefficient in linear power units and σ^0_{dB} is its logarithmic representation. The logarithmic form is convenient for visualisation and analysis because radar backscatter values can span a broad dynamic range.

Backscatter must not be interpreted as a simple one-to-one property of land cover. At a conceptual level, the measured response can be expressed as

$$\sigma^0 = f\left(\lambda, \theta, p, \varepsilon, r, g, v, m\right), \tag{1.3}$$

where λ is the radar wavelength, θ represents acquisition geometry such as incidence angle, p denotes polarisation, ε represents dielectric properties, r represents surface roughness, g represents geometric structure, v represents vegetation structure, and m represents moisture-related conditions. Equation 1.3 is not intended as a physical inversion model; it illustrates that the observed SAR intensity results from several interacting factors rather than from the semantic class label alone [6].

This dependence explains why SAR is informative for land cover. Calm open water commonly produces a different scattering response from rough urban structures; vegetation interacts with microwaves differently from smooth bare surfaces; and structural characteristics within vegetation can change the amount and polarisation of returned energy. The same dependence also explains why classification is difficult. Wind can modify water-surface roughness, precipitation and soil moisture can alter dielectric properties, vegetation changes over time, and topography changes local imaging geometry. A robust classifier must therefore learn patterns that remain useful despite these sources of variability.

Polarisation provides an additional dimension. A radar system can transmit and receive electromagnetic waves with different polarisation orientations. Sentinel-1 land observations are commonly available in VV and VH channels, among other acquisition configurations. Co-polarised and cross-polarised measurements respond differently to surface and volume scattering. In the implemented workflow of this thesis, the Sentinel-1 predictor sequence is constructed from the **VH** polarisation. The exact collection filters, orbit selection, acquisition year, processing steps, and feature generation are documented in Chapter 4 rather than duplicated here.

Time is equally important. Let a sequence of SAR observations for a location or image patch be represented as

$$\mathcal{S} = \{S(t_1), S(t_2), \ldots, S(t_N)\}, \tag{1.4}$$

where $S(t_n)$ is the radar observation acquired at time t_n. A single-image classifier receives only one realisation of the surface response. A multitemporal classifier can instead exploit variation across the sequence. Such variation may provide discriminative information when classes have different temporal behaviour.

Directly presenting a long and irregular image sequence to a model is not always the most effective representation. Acquisition availability can vary in time, the number of scenes can differ geographically, and computational requirements increase as the sequence grows. Temporal aggregation can provide a compromise between single-date mapping and dense time-series modelling. Rather than discarding the temporal dimension entirely, the sequence can be divided into meaningful intervals, and each interval can be represented through a set of synthesised descriptors.

This idea is closely related to the approach developed by Russo et al. for land-cover mapping with Sentinel-1, in which seasonal synthesized spatio-temporal features are coupled with a Swin-Unet architecture [9]. Their work demonstrates that temporally organised Sentinel-1 information can support land-cover classification over geographically distinct environments and motivates the feature strategy adopted and adapted in this thesis.

For the Philippine workflow, all Sentinel-1 acquisitions considered for model input belong to the year **2019**. The sequence is divided into four fixed temporal groups. Seven feature representations are then generated for each group.

$$N_{\text{features}} = N_{\text{groups}} \times N_{\text{features/group}} = 4 \times 7 = 28. \tag{1.5}$$

At this introductory stage, the 28 channels can be written compactly as

$$\mathcal{F} = \bigcup_{q=1}^{4} \{\text{SI}_q, \text{LEE}_q, \text{MAX}_q, \text{MIN}_q, \text{MAXMIN}_q, \text{MEAN}_q, \text{MEDIAN}_q\}, \tag{1.6}$$

where q denotes one of the four temporal groups. Here, SI represents the synthesized Super Image, LEE represents its Lee-filtered form, and the remaining channels encode local statistical information. Equation 1.6 defines only the organisation of the model input. The complete computation of each feature, the precise temporal intervals, the multitemporal denoising procedure, the neighbourhood operations, and the export workflow are presented in Chapter 4.

An important terminological decision is made here. Although the processing script internally uses the labels “winter”, “spring”, “summer”, and “autumn” for the four date intervals, these names are not climatologically appropriate descriptions of Philippine seasons. The Philippines is more naturally described through rainfall regimes and rainy/dry seasonal patterns [11]. We therefore use the neutral

term *temporal group* in the scientific description and preserve the implemented calendar boundaries when the workflow is formally documented in Chapter 4.

The transformation from many SAR acquisitions to 28 standardised feature channels has two purposes. First, it preserves information from different parts of the year without requiring the segmentation network to process an arbitrarily long sequence of individual acquisitions. Second, it exposes the model to complementary forms of information: temporally synthesised intensity, filtered intensity, local extremes, local range, and local central-tendency statistics. Whether these channels are sufficient to separate all classes is an empirical question that must be assessed through the final classification results.

SAR therefore provides the observational foundation of this thesis, but it does not operate independently of the remaining methodology. The sensor provides repeatable measurements; feature extraction transforms those measurements into a model-compatible representation; reference-map generation provides class supervision; and semantic segmentation learns the mapping between both. The value of Sentinel-1 in this work arises from this complete processing chain rather than from any individual image.

1.5 Deep Learning for Semantic Segmentation

Once the reference labels and Sentinel-1 features have been prepared, the land-cover problem can be formulated as semantic segmentation. Semantic segmentation is a dense prediction task in which a model assigns a semantic class to every spatial element of an input image. This differs from image-level classification, where a complete image receives one label, and from object detection, where the output consists primarily of discrete object locations. For land-cover mapping, dense segmentation is a natural formulation because the required output is a spatially continuous categorical map.

Let an input patch be represented by

$$\mathbf{X} \in \mathbb{R}^{C \times H \times W}, \tag{1.7}$$

where H and W are the spatial dimensions and C is the number of input feature channels. In the present workflow,

$$C = 28, \tag{1.8}$$

corresponding to the four temporal groups and seven Sentinel-1-derived features per group introduced in Section 1.4.

The corresponding categorical reference raster is

$$\mathbf{Y} \in \{0, \ldots, K-1\}^{H \times W}, \tag{1.9}$$

where K denotes the number of output codes. In the Philippine implementation, $K = 10$, corresponding to IDs 0–9. IDs 1–9 represent the nine land-cover classes, while ID 0 represents NoData/unclassified pixels and is excluded from the supervised target pixels during optimisation, as described in Chapter 5.

A neural network parameterized by $\boldsymbol{\theta}$ learns a mapping

$$f_{\boldsymbol{\theta}} : \mathbb{R}^{C \times H \times W} \to \mathbb{R}^{K \times H \times W}. \tag{1.10}$$

For every spatial location (i, j), the network produces one logit for each class. These scores can be converted to class probabilities using the softmax function,

$$p_{i,j,k} = \frac{\exp(z_{i,j,k})}{\sum_{r=0}^{K-1} \exp(z_{i,j,r})}, \qquad k = 0, \ldots, K-1, \tag{1.11}$$

and the pixel-wise prediction is

$$\widehat{y}_{i,j} = \arg\max_{k} p_{i,j,k}. \tag{1.12}$$

These equations describe the general semantic-segmentation task. The exact loss function, optimisation algorithm, training schedule, batch configuration, class-balancing strategy, and evaluation metrics used in the final experiments are intentionally reserved for Chapter 5, where they can be documented together with the final implementation.

Deep learning is particularly attractive for this task because land-cover classification is not determined only by an individual pixel value. Spatial context matters. A group of bright radar responses arranged in a structured pattern can carry different semantic meaning from identical values occurring as isolated points. Forest, agriculture, water, built-up surfaces, and other classes also exhibit characteristic spatial textures and neighbourhood relationships. A deep segmentation model learns hierarchical representations that combine local measurements with progressively broader spatial context.

The development of fully convolutional networks was an important step in modern dense prediction. Long et al. showed that classification networks could be reformulated as fully convolutional architectures that receive an image and produce correspondingly spatial outputs, enabling efficient end-to-end pixel-level prediction [12]. This formulation provided a foundation for many subsequent segmentation networks.

U-Net introduced an encoder–decoder structure with skip connections between corresponding spatial scales [13]. The encoder progressively transforms the input into increasingly abstract feature representations while reducing spatial resolution. The decoder progressively recovers spatial resolution to generate a dense output. Skip connections transfer higher-resolution information from encoder stages to corresponding decoder stages, helping the model recover boundaries and spatial details that can otherwise be lost during downsampling.

In a simplified form, an encoder–decoder segmentation process can be written as

$$\mathbf{E}_\ell = \mathcal{E}_\ell \left(\mathbf{E}_{\ell-1} \right), \tag{1.13}$$

$$\mathbf{D}_{\ell-1} = \mathcal{D}_{\ell-1} \left(\mathbf{D}_\ell, \mathbf{E}_{\ell-1} \right), \tag{1.14}$$

where $\mathbf{E}_\ell$ represents an encoder feature map at level ℓ, $\mathbf{D}_\ell$ represents a decoder feature map, and $\mathcal{E}_\ell$ and $\mathcal{D}_\ell$ denote learned transformations. The presence of $\mathbf{E}_{\ell-1}$ in Equation 1.14 represents the conceptual role of a skip connection: the decoder can combine high-level semantic information with features from a finer spatial scale.

Convolutional architectures are effective at learning local spatial patterns, but their receptive fields and information propagation are determined by the organisation and depth of convolutional operations. Transformer architectures provide a different mechanism based on self-attention, allowing feature interactions to be modelled through learned relationships between tokens. Applying standard global self-attention directly to high-resolution imagery is computationally expensive because the number of pairwise token interactions grows rapidly with image size.

The Swin Transformer addresses this problem through hierarchical representations and shifted-window self-attention [14]. Rather than computing attention globally across all image tokens at every layer, attention is calculated within local windows. Alternating shifted windows allow information to propagate across window boundaries while preserving a computational structure that is better suited to high-resolution vision tasks. The hierarchical design also produces multiscale representations, which are useful for dense prediction.

For an attention operation with query, key, and value matrices $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$, the general scaled dot-product attention operation is

$$\text{Attention} \left(\mathbf{Q}, \mathbf{K}, \mathbf{V} \right) = \text{Softmax} \left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}} \right) \mathbf{V}, \tag{1.15}$$

where d is the dimensionality used to scale the query–key product. In a Swin Transformer, this mechanism is applied within spatial windows and is combined with the shifted-window strategy [14]. Chapter 5 provides the architecture-specific

description required for the model used in this thesis. Swin-Transformer blocks can be incorporated into U-shaped encoder–decoder segmentation architectures so that hierarchical self-attention is combined with multiscale dense prediction. This methodological family is especially relevant to the present project because Russo et al. applied a Swin-Unet-based framework to land-cover mapping from seasonal Sentinel-1 features [9]. Their experiments over Amazonian, African, and Siberian environments demonstrated the potential of combining synthesised temporal SAR representations with transformer-based semantic segmentation. The present thesis follows this methodological direction while implementing a two-dimensional SwinUNETR configuration in MONAI [15, 16], as described in detail in Chapter 5.

The present thesis builds on this methodological direction but addresses a different geographic setting and a different training-reference construction problem. The Philippines introduces tropical climatic conditions, extensive coastal environments, diverse vegetation and agricultural patterns, heterogeneous settlements, and a reference dataset created through a Philippines-specific extension of the MOLCA agreement concept. As a result, model performance observed in previous geographic regions cannot be assumed in advance. The Philippine experiment provides a new test of how the spatio-temporal Sentinel-1 and Swin-Unet methodology behaves when paired with an adapted agreement-derived training dataset.

Recent remote-sensing literature supports the broader use of semantic segmentation for land-cover mapping. A systematic review by Ajibola and Cabral found extensive use of encoder–decoder and related deep architectures in land-cover applications and identified Sentinel imagery among the prominent sources used in this research area [4]. Studies have also demonstrated the application of U-Net to combined SAR and multispectral land-cover classification [17], while evaluations of spatio-temporal architectures show that models capable of preserving temporal structure, including vision transformers, can effectively learn from Sentinel time series [8].

The main reason for using a deep segmentation model in this thesis is therefore not simply that deep learning is more recent than conventional classifiers. The methodological motivation is more specific. The Sentinel-1 input is multichannel and spatio-temporal; the output is a dense categorical map; land-cover interpretation benefits from neighbourhood context; class boundaries occur at multiple spatial scales; and the training dataset is large enough to motivate a model capable of learning complex representations. A Swin-Unet-based segmentation framework is well aligned with these characteristics.

At the same time, model complexity does not remove the need for careful evaluation. Deep networks can reproduce systematic label errors, favour dominant classes, or learn geographic correlations that fail outside the training distribution. Overall accuracy alone can hide poor performance in minority classes. For that

reason, the final model must be evaluated using class-specific as well as aggregate metrics, and the confusion matrix must be examined to identify which classes are being confused. These procedures are developed in Chapters 5 and 6.

Figure 1.4 provides a simplified view of the model-level task used in the thesis. The complete architecture contains many internal transformations; the figure is intended only to clarify the relationship between the 28-channel predictor tensor and the final pixel-level land-cover map.

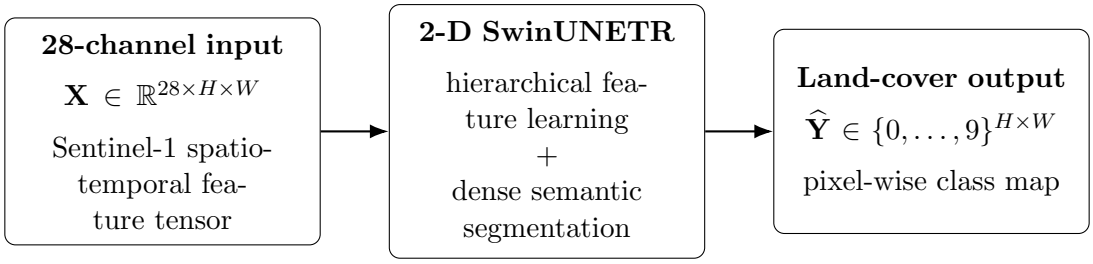


Figure 1.4: Simplified semantic-segmentation formulation used in the Philippine experiments. Each spatial sample is represented by 28 Sentinel-1-derived feature channels and processed by the 2-D SwinUNETR model to produce a ten-code pixel-wise output comprising NoData/unclassified (ID 0) and nine land-cover classes (IDs 1–9).

1.6 Aim and Objectives of the Thesis

The primary aim of this thesis is to develop and evaluate a high-resolution land-cover mapping workflow for the Philippines in which a Philippines-specific MOLCA-derived reference dataset provides training labels and multitemporal Sentinel-1 SAR features provide the input to a deep semantic-segmentation model.

More specifically, the study investigates whether the agreement-based reference-data philosophy introduced by MOLCA [3] can be adapted to the Philippine context and combined with the Sentinel-1 spatio-temporal deep-learning strategy developed in recent work [9]. The thesis therefore joins two methodological components that are often treated separately: the construction of reliable large-scale labels and the learning of land-cover information from multitemporal SAR imagery.

To achieve this aim, the thesis addresses the following objectives:

1. **Construct a Philippines-specific MOLCA-derived land-cover reference dataset.**

Multiple existing land-cover products are collected and prepared for the Philippine study area. Their spatial characteristics and thematic legends are harmonised so that agreement can be evaluated on a common basis. The workflow begins from the MOLCA philosophy but is adapted to the datasets and thematic requirements of the Philippines.

2. **Integrate global, thematic, national, and regional land-cover information within a common agreement framework.**

   The reference-generation workflow uses eight source datasets with different roles and characteristics. In addition to products inherited from the general MOLCA logic, Philippines-relevant regional information is incorporated. The adaptation also includes specific thematic refinements required for the study area, including the treatment of mangrove-related land cover. The complete datasets and agreement rules are documented in Chapter 2 and 3.

3. **Harmonize the spatial and thematic properties of the reference products.**

   Differences in coordinate reference system, spatial resolution, raster alignment, extent, class coding, and legend semantics must be resolved before agreement is calculated. This objective ensures that agreement represents thematic correspondence rather than artefacts caused by incompatible raster geometry.

4. **Develop a systematic geographic sampling strategy for deep learning.**

   The study uses the Sentinel-2 tiling grid as a spatial organisational framework. Sixty Sentinel-2 tiles are selected to provide geographic coverage, after which candidate training areas are evaluated according to class availability and spatial distribution. The resulting procedure produces 280 selected reference patches across the 60 operational tiles. During the subsequent Sentinel-1 availability check, three selected patches from tile `51QUC` could not provide the complete required 2019 VH feature set. The final deep learning dataset therefore contains 277 valid scene–reference associations.

5. **Prepare a multitemporal Sentinel-1 SAR dataset for the selected Philippine areas.**

   Sentinel-1 observations from 2019 are collected using a reproducible processing workflow. The implemented configuration uses Interferometric Wide Swath data, descending acquisitions, and VH polarisation. The acquisition sequence is organised spatially according to the selected tiles and patches.

6. **Transform the Sentinel-1 time series into a compact spatio-temporal feature representation.**

   The annual Sentinel-1 sequence is divided into four fixed temporal groups. For each group, seven feature channels are generated: the Super Image, a Lee-filtered Super Image, local maximum, local minimum, local maximum–minimum range, local mean, and local median. The resulting input therefore consists of 28 feature channels per patch. The complete mathematical and computational derivation is provided in Chapter 4.

7. **Apply deep semantic segmentation to the Sentinel-1 feature representation.**

   A two-dimensional SwinUNETR model implemented in MONAI is used to learn the relationship between the 28-channel Sentinel-1 input and the MOLCA-derived land-cover labels. The architecture combines a hierarchical Swin Transformer encoder with a U-shaped segmentation framework and multi-scale skip connections. Model design and training are described in Chapter 5.

8. **Evaluate classification performance at both overall and class-specific levels.**

   The resulting predictions are evaluated using confusion matrices, Overall Accuracy, class-wise precision, recall, F1-score and Intersection over Union (IoU), together with macro-averaged measures, mean IoU, and Cohen's kappa. Particular attention is given to differences among classes so that strong aggregate performance cannot conceal weak behaviour for less represented or more difficult land-cover categories.

9. **Analyse limitations and sources of classification error.**

   The study examines how reference-data quality, class representation, geographic variability, SAR characteristics, temporal feature construction, and model behaviour influence the final results. Difficult class pairs and spatial patterns of error are considered explicitly in the discussion.

10. **Establish a reproducible workflow that can support future extension.**

    Dataset preparation, geographic sampling, Sentinel-1 processing, feature generation, and model-ready data organisation are implemented through systematic processing steps and scripts. This allows the study to be reproduced and provides a basis for extension to additional Philippine areas, alternative temporal periods, or future model configurations.

These objectives define the thesis as more than an isolated neural-network experiment. The contribution lies in the complete path from heterogeneous existing land-cover information to a harmonised agreement map, from the agreement map to geographically organised training patches, from raw multitemporal radar observations to a standardised feature tensor, and from that tensor to a dense land-cover prediction.

The work also separates the roles of *reference information* and *predictor information*. The global, thematic, national, and regional land-cover products are used to construct the MOLCA-derived labels. They are not the model's primary remote-sensing predictor channels. Sentinel-1 SAR provides those predictor channels. Maintaining this distinction is important when interpreting the final model: the segmentation network is trained to learn a relationship between SAR-derived features and agreement-derived land-cover labels.

The study does not assume that the agreement-derived dataset constitutes perfect ground truth. Agreement among multiple products is used as a practical mechanism for increasing confidence and providing sufficient training information at high spatial resolution. Residual uncertainties can remain because the source products have their own errors, baseline years, class definitions, and spatial properties. The quality of the final classification must therefore be interpreted in relation to both the model and the reference-generation procedure.

Likewise, the thesis does not claim that 28 handcrafted Sentinel-1 features represent the only possible SAR input configuration. They constitute the implemented feature representation selected for this experiment and are motivated by the use of temporal synthesis and spatial descriptors in the related Sentinel-1 deep-learning methodology [9]. Their practical value is assessed through the resulting classification performance.

The central research workflow can therefore be summarised by the following composition:

$$\begin{aligned} &\underbrace{\text{existing LC products}}_{\text{reference sources}} \longrightarrow \underbrace{\text{MOLCA-derived labels}}_{\text{training supervision}} \\ &\underbrace{\text{2019 Sentinel-1 sequence}}_{\text{EO observations}} \longrightarrow \underbrace{\text{28 SAR features}}_{\text{model input}} \\ &\longrightarrow \underbrace{\text{2-D SwinUNETR}}_{\text{semantic segmentation}} \longrightarrow \underbrace{\text{land-cover map}}_{\text{prediction}}. \end{aligned} \tag{1.16}$$

Equation 1.16 provides the concise conceptual definition of the thesis. The following chapters decompose each operation and document the data, algorithms,

implementation decisions, and results required to reproduce it.

1.7 Thesis Structure

The remainder of the thesis is organised so that the construction of the reference dataset is separated from the selection of training areas, Sentinel-1 feature generation, and deep-learning classification. This separation reflects the fact that reference-map generation and SAR-based classification solve different methodological problems and require different forms of validation and documentation.

Chapter 2, Study Area and Datasets, introduces the geographical context of the Philippines and describes the datasets used throughout the study. It presents the Sentinel-2 tiling grid adopted as the spatial reference framework, the global land-cover products, the regional land-cover products, and the Sentinel-1 SAR dataset. The chapter also documents the relevant coordinate reference systems, spatial resolutions, baseline years, and other characteristics required to understand the later harmonisation and processing steps. Its purpose is to describe what data are used before explaining how those data are transformed.

Chapter 3, Generation of the MOLCA Philippines Reference Map, focuses exclusively on the construction of the reference information. The chapter first introduces the MOLCA concept and identifies the land-cover products participating in the Philippine implementation. It then describes preprocessing, target-legend construction, thematic reclassification, spatial alignment, resampling, and the agreement rules used to create the final MOLCA-derived reference map. Particular attention is given to the differences between the original MOLCA methodology and the adaptations introduced for the Philippines. The final reference dataset and its thematic organisation are documented at the end of the chapter.

Chapter 4, Patch Selection and Sentinel-1 Feature Preparation, moves from reference-map construction to the generation of model-ready training data. Candidate areas are identified using the Sentinel-2 tiling framework, and the strategy used to select the 60 tiles and 280 reference patches is presented, together with the Sentinel-1 availability verification that results in 277 valid samples for the deep learning stage. Class and geographic distributions are analysed to evaluate the representativeness of the selected areas. The chapter then describes the collection and preprocessing of 2019 Sentinel-1 data, the formation of four temporal groups, the multitemporal processing procedure, and the complete mathematical construction of the 28 Sentinel-1 features. Finally, it explains how the SAR feature rasters are aligned with the reference patches.

Chapter 5, Deep Learning Classification Methodology, describes the semantic-segmentation stage. It introduces the final model architecture and explains the Swin Transformer and U-Net components required to understand the

implemented SwinUNETR configuration. The organisation of the 28 input features and output classes is specified, followed by the training, validation, and final evaluation strategy. The chapter also presents the loss function, optimiser, hyperparameters, class-balancing procedures, and accuracy metrics used in the final experiments.

Chapter 6, Results and Discussion, reports and interprets the results produced by the complete workflow. It first summarises the Philippines MOLCA-derived reference dataset and selected-patch dataset, followed by the generated Sentinel-1 feature dataset. Deep-learning training behaviour and final segmentation results are then presented using quantitative metrics, confusion matrices, class-wise statistics, and spatial examples. The discussion examines successful and difficult classes, sources of error, effects of data imbalance and sampling, limitations of the reference dataset, and the implications of using SAR-only predictors for high-resolution tropical land-cover mapping.

Chapter 7, Conclusions, synthesizes the principal findings of the research. It revisits the original objectives, identifies the main methodological and experimental conclusions, discusses the principal limitations, and outlines directions for future work. Particular attention is given to the potential extension of the agreement-derived training strategy and Sentinel-1 deep-learning workflow to larger areas or future land-cover mapping experiments.

The thesis is completed by a set of **Appendices**. These contain supporting technical material that is important for reproducibility but would interrupt the narrative if included in full within the main chapters. The appendices include the detailed land-cover legend-harmonisation tables, the list of selected Sentinel-2 tiles, patch-selection statistics, relevant Python and Google Earth Engine processing scripts, and additional maps, figures, tables, and model results.

Finally, a unified **Bibliography** contains all scientific publications, official dataset documentation, mission references, and other sources cited throughout the thesis. References are managed centrally in the Overleaf project so that each source is stored only once and citations are automatically numbered according to their first appearance in the thesis.

Together, these chapters follow the complete research sequence from motivation to data, from reference construction to predictor preparation, from deep-learning methodology to experimental evaluation, and finally to the interpretation and conclusions of the work.

Chapter 2

Study Area and Datasets

2.1 Study Area: The Philippines

Land-cover mapping starts from a clear definition of the geographical area to be mapped. In this thesis, the study area is the Philippines, an archipelagic country in Southeast Asia composed of thousands of islands distributed between the South China Sea, the Philippine Sea, the Sulu Sea, and the Celebes Sea. From the point of view of Earth Observation, this geographical configuration is important because it combines several conditions that make high-resolution land-cover mapping both necessary and technically demanding: fragmented coastlines, narrow islands, mountainous terrain, dense tropical vegetation, highly urbanised areas, agricultural plains, wetlands, mangrove belts, and large inland and coastal water surfaces.

The Philippines is particularly suitable for a land-cover mapping experiment based on radar remote sensing and a consensus-derived reference map. The territory contains land-cover classes that are spectrally and spatially heterogeneous, and some of these classes are difficult to represent consistently across global products. Forest and cropland can dominate large portions of the landscape, while built-up areas may appear as compact metropolitan zones, linear settlements, or small scattered villages. Grassland, shrubland, wetland, and bareland are more spatially fragmented and can be sensitive to the definition adopted by each input land-cover product. Mangroves require particular attention because they are ecologically important coastal forested wetlands, but they can be classified differently depending on whether a dataset prioritises vegetation structure, wetland character, coastal habitat, or national land-cover terminology.

For this reason, the study area was not treated simply as a bounding box. Instead, the project used the Philippine national extent together with the Sentinel-2 tiling grid as the operational spatial framework. The use of tiles allowed the work

to be divided into computationally manageable units, which was important because the full workflow required clipping, harmonising, reclassifying, and aligning several large raster datasets. This tile-based approach also made the processing reproducible: the same sequence of operations could be tested on a pilot tile, verified, and then repeated for the remaining tiles covering the area of interest.

The administrative boundary used in the study-area figures was obtained from the GADM database, version 4.1 [18]. GADM provides standardised administrative boundary data for countries worldwide and distributes the datasets in geospatial formats including GeoPackage and shapefile. In this thesis, the Philippine boundary is used only as a cartographic support layer for representing the study area and its relationship with the Sentinel-2 tiling framework; it does not contribute any land-cover labels to the reference-generation procedure.

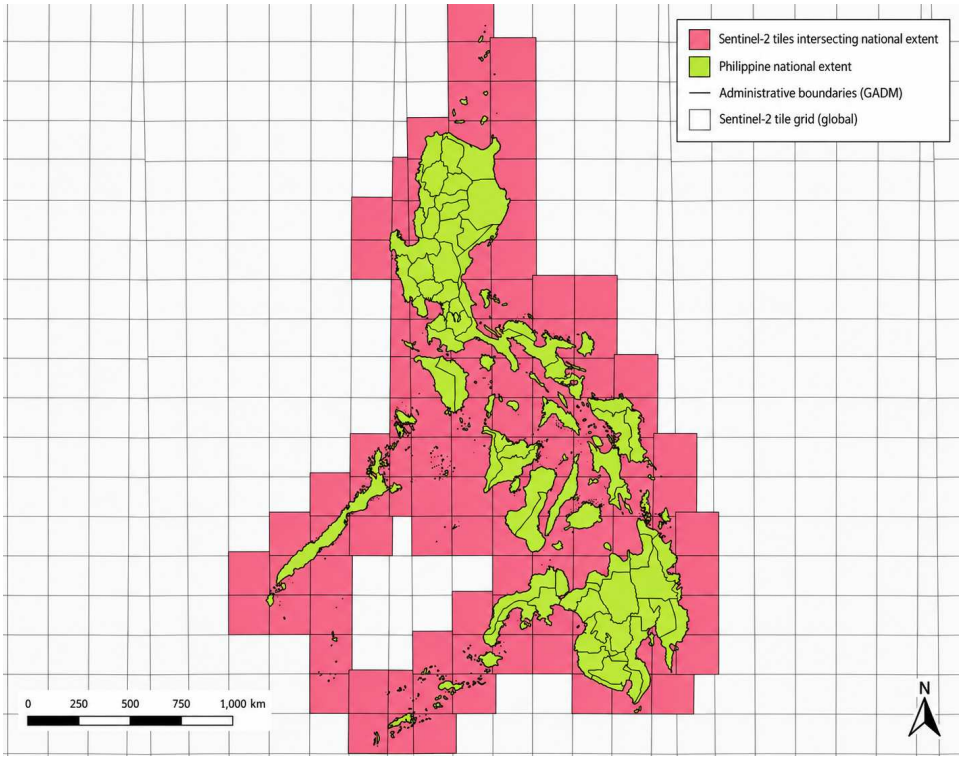


Figure 2.1: Philippine study area and Sentinel-2 tile coverage intersecting the national extent. The GADM administrative polygons provide geographical context, while the Sentinel-2 grid defines the spatial framework used for tile-based processing.

The country-wide setting is relevant for the design of the thesis because the final land-cover mapping problem is not limited to one urban area or one ecological zone. The selected area includes northern Luzon, central island groups, Mindanao, and

coastal/island environments. This diversity increases the value of a reference map generated from multiple sources, because a single global product may perform well in one type of landscape but less reliably in another. A consensus-based reference is therefore useful not only to increase label reliability, but also to reduce dependence on the specific errors of one input dataset.

2.2 Sentinel-2 Tiling Grid Used as Spatial Reference

Although the final classification model uses Sentinel-1 SAR features, the spatial organisation of the thesis is based on the Sentinel-2 tiling system. This choice follows the logic used in previous high-resolution land-cover mapping workflows, including the original MOLCA framework and the Sentinel-1/Swin-Unet land-cover architecture developed at the University of Pavia [3, 9]. Sentinel-2 Level-1C and Level-2A products are distributed in tiles defined in a UTM/WGS84 cartographic reference frame. The Sentinel-2 tiling grid follows the 100 km MGRS grid spacing, while individual distributed product tiles cover approximately 110 km $\times$ 110 km to provide overlap between neighbouring tiles. Each tile is identified by a five-character MGRS tile code [19, 20].

Using the Sentinel-2 grid has three main advantages for this thesis. First, it provides a standard spatial indexing system. Each tile can be processed, checked, and referred to by a unique identifier such as 51PUS or 51PTR. Second, it avoids processing the entire Philippines in a single operation, which would be inefficient for large raster inputs. Third, it creates a direct bridge between the MOLCA-derived reference map, the selected training patches, and the Sentinel-1 features generated later in the workflow.

The operational area was first obtained by intersecting the Sentinel-2 global tiling grid with the Philippines area of interest. This produced the set of tiles overlapping the country. From this broader set, a subset of 60 Sentinel-2 tiles was retained for the main project workflow. The detailed selection strategy is discussed later in Chapter 4, because it belongs to patch selection and feature preparation. In this chapter, the important point is that the Sentinel-2 tiling grid defines the spatial reference units used to organise the datasets.

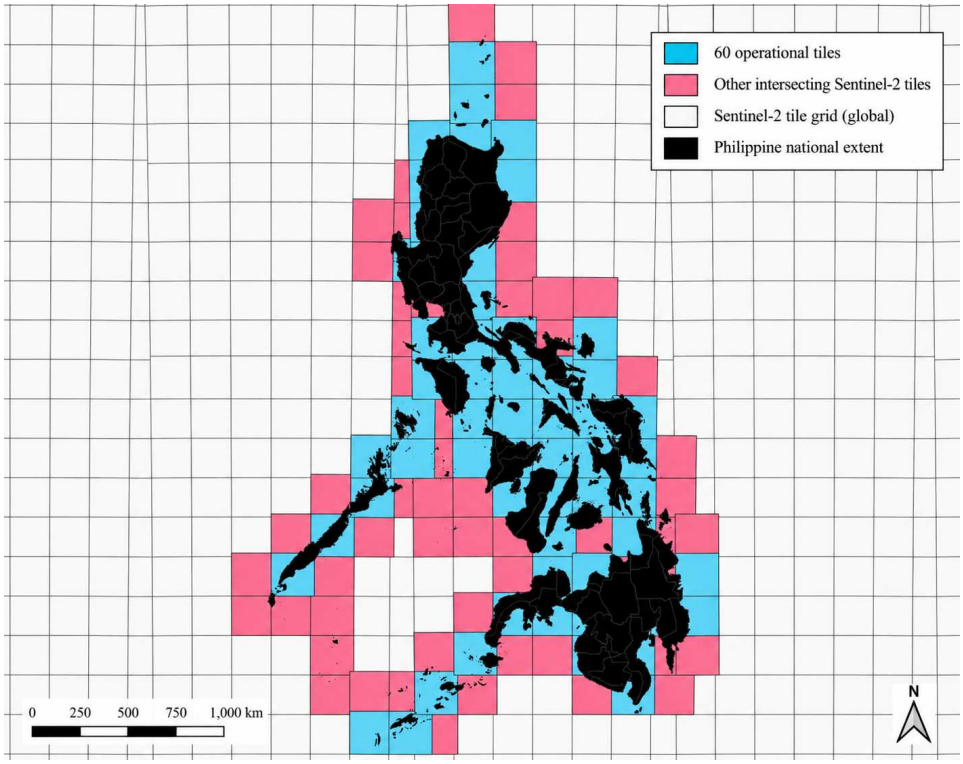


Figure 2.2: Sentinel-2 tiling framework used in the Philippine workflow. The 60 operational tiles retained for MOLCA processing and subsequent Sentinel-1 feature preparation are highlighted in cyan, while the additional tiles intersecting the broader Philippine extent are shown in pink.

Table 2.1 summarises the spatial framework adopted in the thesis. The table is intentionally placed before the dataset descriptions because it explains how all later data layers are spatially organised.

Table 2.1: Spatial framework adopted for the Philippine land-cover mapping workflow.

Component	**Role in the thesis**
Study area	Philippines, represented by national administrative boundary data and processed through the Sentinel-2 tile framework.
Boundary layer	GADM Philippines administrative boundary, used for study-area representation and map layout.
Spatial indexing system	Sentinel-2 MGRS tiling grid, used to divide the national area into reproducible operational units.
Operational tile subset	60 selected Sentinel-2 tiles used for the main MOLCA-derived reference map and Sentinel-1 feature workflow.
Tile-level processing logic	Each tile can be clipped, harmonised, reclassified, aligned, and checked independently before scaling to other tiles.
Later patch framework	The 549 × 549-pixel patches are derived inside the selected tile framework and are described in Chapter 4.

2.3 Input Products for the MOLCA-Derived Reference

The deep learning component of this thesis requires a reference dataset that can be used to train and evaluate a semantic segmentation model. Creating such a reference is often one of the most difficult parts of large-area land-cover mapping. Manual labelling over a country-scale area is expensive and time-consuming, while relying on a single existing land-cover product can transfer the errors and class definitions of that product directly into the model.

The MOLCA approach addresses this problem by combining existing high-resolution land-cover products and retaining areas where multiple datasets satisfy predefined agreement conditions [3]. The original MOLCA study generated approximately 117 billion reference pixels at 10 m resolution across large regions of Sub-Saharan Africa, the Amazon, and Siberia. Its published validation reported an Overall Accuracy of approximately 96% [3]. This value refers to the validation performed in the original MOLCA study and must not be interpreted as an accuracy estimate for the Philippine MOLCA-derived reference developed in this thesis.

For the Philippines, the same logic was adapted to the available global and regional datasets. The initial dataset group followed the original MOLCA structure:

two general land-cover products were used as the principal land-cover sources, and thematic products were used to reinforce classes such as built-up, water, wetland, and forest. Later, two regional products were introduced to better represent Philippine landscapes and to reduce the area of no agreement, especially where the global products were insufficient or ambiguous. The detailed rule hierarchy is described in Chapter 3; this chapter focuses on what the datasets are, why they are relevant, and what role they play.

2.3.1 General-purpose global land-cover products

FROM-GLC 2017

FROM-GLC, the Finer Resolution Observation and Monitoring of Global Land Cover, is a global land-cover mapping initiative developed by Tsinghua University. The 2017 10 m global product was generated using Sentinel-2 imagery with a stable classification strategy and a transferred training sample set [21]. In the original MOLCA work, FROM-GLC 2017 is one of the main general-purpose global land-cover inputs [3].

In this thesis, FROM-GLC has two roles. First, it is one of the two principal global land-cover maps used in the base agreement. Second, it provides the spatial reference grid for the Philippine implementation. This is important because the datasets used in the workflow have different native resolutions, coordinate systems, extents, and tiling structures. Using one raster as a reference grid makes it possible to align the other inputs consistently before agreement computation. The use of FROM-GLC as the reference grid also reflects the practical advantage of its 10 m resolution, which matches the target resolution of the MOLCA-derived reference map.

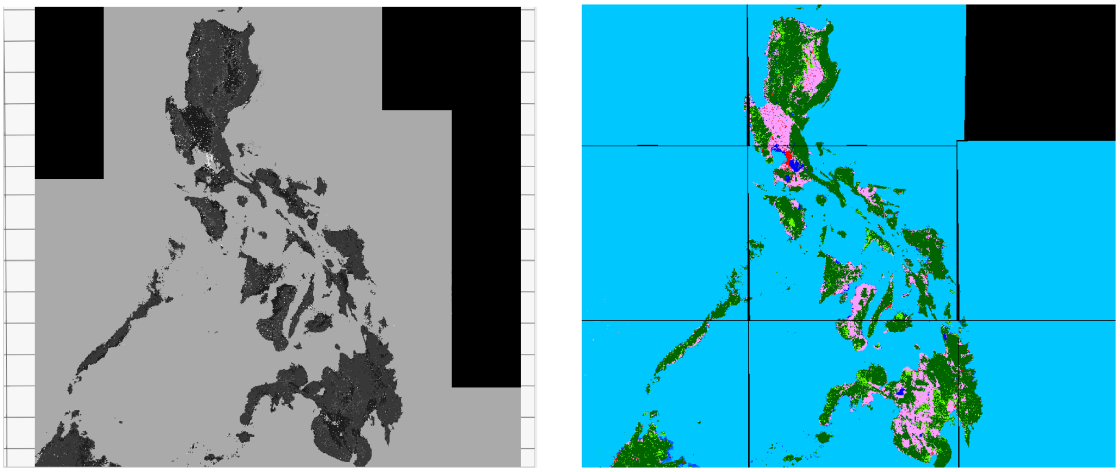
The FROM-GLC legend includes land-cover classes that correspond well to the main classes required in this thesis, including built-up, cropland, forest, grassland, shrubland, wetland, water, and bareland. Nevertheless, semantic differences still exist. For example, a class named "shrub" in one dataset may not exactly match the definition of shrubland in another. These semantic issues are not solved in Chapter 2; they are treated formally through legend harmonisation and reclassification in Chapter 3.

GlobeLand30 2020

GlobeLand30 is a global land-cover product produced by the National Geomatics Center of China. It provides global coverage at 30 m resolution and includes a 10-class land-cover legend. The GlobeLand30 production approach is described as an operational global mapping method based on pixel-object-knowledge integration

[22]. In the original MOLCA paper, GlobeLand30 is used as a general-purpose global land-cover product and is reported as having 30 m resolution with UTM-based distribution [3].

In this thesis, GlobeLand30 2020 is used as the second general-purpose global land-cover product. Together with FROM-GLC, it forms the base agreement stage of the MOLCA-derived reference. The difference between the 10 m resolution of FROM-GLC and the 30 m resolution of GlobeLand30 is not a minor technical detail. It means that GlobeLand30 must be aligned to the reference grid before pixel-level comparison. It also means that some narrow landscape elements may be represented differently between the two products. This can reduce agreement in fragmented Philippine landscapes, but it also makes retained agreement areas more conservative.



(a) FROM-GLC 2017. (b) GlobeLand30 2020.

Figure 2.3: Master coverage views of the two general-purpose global land-cover products used to construct the base agreement of the Philippine MOLCA-derived reference.

2.3.2 Thematic global evidence products

The original MOLCA methodology does not rely only on general-purpose land-cover products. It also uses thematic products that are designed to represent particular classes with higher confidence. This is scientifically useful because thematic products are often produced with class-specific algorithms, class-specific validation, or dedicated sensor choices. In the Philippine workflow, four thematic global products were used: GHS-BUILT-S E2018, WSF 2019, JRC Global Surface Water Yearly History 2019, and JAXA FNF 2019.

GHS-BUILT-S E2018

The Global Human Settlement Layer (GHSL), produced by the Joint Research Centre of the European Commission, provides several global datasets describing the spatial distribution of human settlements. The built-up evidence layer used in the Philippine workflow is the GHS-BUILT-S R2023A product for epoch 2018 [23].

The 10 m version used in the project is distributed in the World Mollweide coordinate reference system (EPSG:54009). Its raster values are stored as unsigned 8-bit integers from 0 to 100 and represent the amount of built-up surface, in square metres, contained within each 10 m cell; value 255 is reserved for NoData [23].

In this thesis, GHS-BUILT-S is not treated as a complete land-cover map. Its role is class-specific: it provides additional evidence for the Built-up class where the two general-purpose land-cover products do not produce a valid base agreement. This makes it complementary to the World Settlement Footprint, which provides a second independent settlement-related evidence source. The exact source-specific reclassification and its position within the MOLCA agreement hierarchy are described in Chapter 3.

World Settlement Footprint 2019

The World Settlement Footprint 2019 (WSF 2019), produced by the German Aerospace Center (DLR), is a 10 m global binary settlement mask derived from multitemporal Sentinel-1 and Sentinel-2 observations acquired during 2019 [24]. It represents the spatial extent of human settlements rather than a complete multi-class land-cover legend.

In the Philippine workflow, WSF 2019 is used together with GHS-BUILT-S as a second source of Built-up evidence. The two products differ in their underlying processing strategy and representation of settlement information, which makes them complementary within the agreement-based reference construction.

The source-specific raster coding and the exact reclassification rule applied to WSF are documented in Chapter 3. Chapter 2 therefore treats WSF only at the dataset and functional-role level.

JRC Global Surface Water Yearly History

The Global Surface Water dataset developed by the Joint Research Centre provides long-term information on the spatial and temporal distribution of surface water from the Landsat archive [25]. Several derived Global Surface Water products are available, including occurrence, seasonality, transitions, and yearly water-history layers.

During the initial dataset-preparation stage, the occurrence product was examined. The final Philippine MOLCA workflow, however, uses the *Yearly History* product for 2019 because its categorical `waterClass` representation distinguishes non-water, seasonal water, and permanent water explicitly [26].

The Yearly History coding used in the final workflow is:

- 0: NoData;
- 1: Not water;
- 2: Seasonal water;
- 3: Permanent water.

In the MOLCA Philippines workflow, this product is used as a thematic evidence source rather than as a general-purpose land-cover classification. Permanent-water pixels support the Water class, while seasonal-water pixels provide evidence for the Wetland class. The exact reclassification and its position within the agreement hierarchy are described in Chapter 3.

JAXA Forest/Non-Forest product

The Forest/Non-Forest (FNF) product developed by the Japan Aerospace Exploration Agency (JAXA) is a thematic global dataset derived from L-band Synthetic Aperture Radar (SAR) observations acquired by the ALOS PALSAR and ALOS-2 PALSAR-2 sensors. The original global FNF maps were produced from annual PALSAR mosaics at approximately 25 m spatial resolution, using radar backscatter information to distinguish forest from non-forest areas [27, 28]. The FNF product is available for multiple observation years, and the 2019 FNF map was selected for the MOLCA workflow [3].

In the Philippine workflow, the 2019 FNF layer is used primarily as thematic evidence for the Forest class. Its radar-based origin provides information that is complementary to the optical land-cover products used in the workflow. L-band SAR is particularly useful for forest applications because its backscatter is sensitive to vegetation structure and can be acquired independently of cloud cover [27, 28]. This is relevant in the Philippines, where persistent cloud cover can limit the availability of consistent optical observations.

As a thematic product, FNF does not provide the complete set of land-cover classes required by the target legend. It therefore cannot independently distinguish classes such as Cropland, Grassland, Shrubland, Wetland, or Built-up. In this thesis, FNF is consequently used as class-specific evidence rather than as a complete land-cover map. Before being introduced into the agreement hierarchy,

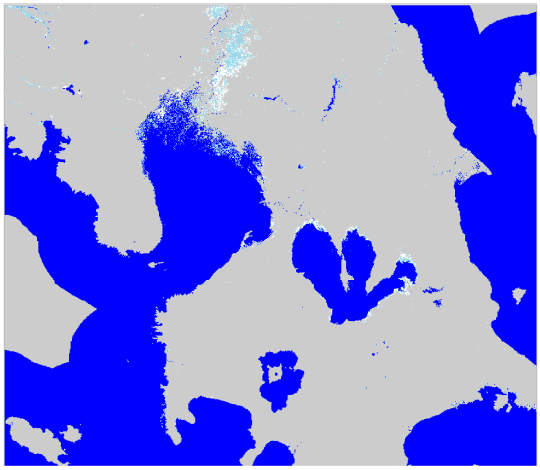
its source classes are translated to the common MOLCA-derived legend. The exact reclassification rules and the position of FNF within the agreement hierarchy are described in Chapter 3.



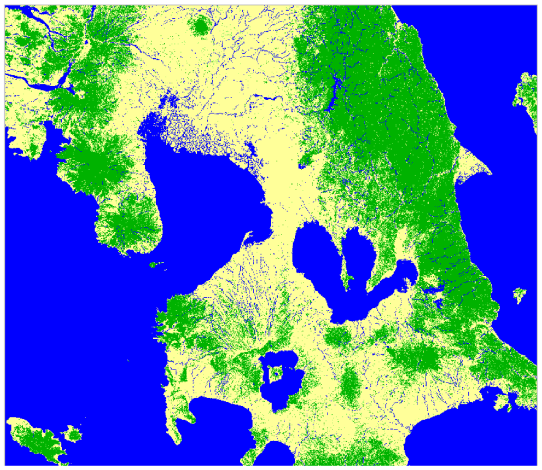
(a) GHS-BUILT-S E2018.



(b) WSF 2019.



(c) JRC Global Surface Water.



(d) JAXA Forest/Non-Forest.

Figure 2.4: Thematic evidence products over a common example area around Manila. The panels illustrate the different spatial and thematic information contributed by the Built-up, surface-water, and forest evidence sources. Each panel uses source-specific symbology; colours and grayscale values should therefore not be interpreted as equivalent class codes across products.

Table 2.2: Global input products used in the MOLCA-derived Philippine reference workflow.

Dataset	Producer	Year	Res.	Product type	Role in this thesis
FROM-GLC	Tsinghua Univ.	2017	10 m	Global land-cover map	Principal global land-cover product and reference grid for agreement.
Globe-Land30	NGCC, China	2020	30 m	Global land-cover map	Second general land-cover product used in the agreement procedure.
GHS-BUILT-S	EC JRC	2018	10 m	Built-up surface	Built-up evidence where general land-cover products do not provide sufficient agreement.
WSF	DLR	2019	10 m	Settlement mask	Complementary settlement and built-up evidence used together with GHS-BUILT-S.
GSW	EC JRC	2019	30 m	Surface-water product	Water and wetland-related evidence in the reference map (Yearly History).
FNF	JAXA	2019	25 m	Forest/non-forest map	Forest evidence used in the agreement hierarchy.

2.4 Regional Land-Cover Products

Global products are essential for large-area land-cover mapping because they provide consistent coverage across countries and continents. However, global consistency can come at the cost of local detail. A global classifier must work across many biomes, terrain types, sensors, and national landscapes. For the Philippines, this means that classes such as mangrove forest, fishpond, brush/shrubs, rubber or oil-palm plantations, and coastal wetlands may not be represented with the local detail required for a reliable training reference.

For this reason, the Philippine workflow was extended with two regional products: the NAMRIA 2020 land-cover map and the JAXA High-Resolution Land-Use and Land-Cover Map of Southeast Asia 2023. This follows the spirit of the original MOLCA paper, where regional datasets were used to strengthen the consensus procedure in specific macro-regions [3]. In the Philippine case, regional products were particularly important because they introduced local/national class knowledge and a Southeast Asian classification product designed for landscapes more similar to the study area.

2.4.1 NAMRIA 2020 Land Cover Map

NAMRIA, the National Mapping and Resource Information Authority, is the central mapping agency of the Government of the Philippines. For this thesis, the NAMRIA 2020 land-cover map is treated as the national land-cover product used to complement the global datasets. Philippine official and statistical sources report 2020 land-cover categories including closed forest, open forest, mangrove forest, brush/shrubs, open/barren, grassland, marshland/swamp, annual crop, perennial crop, fishpond, built-up, inland water, and sea/ocean [29, 30]. These categories are highly relevant to this thesis because they correspond directly to the final MOLCA-derived legend or to classes that must be harmonised into it.

The value of NAMRIA in this thesis is not only that it is national. Its value is that it provides a Philippine-specific interpretation of land cover. A global product may identify broad vegetation patterns, but a national product can use a class system that reflects local land administration, environmental reporting, and national mapping needs. For example, NAMRIA distinguishes closed forest, open forest, and mangrove forest. It also separates annual crops from perennial crops, and includes fishpond as a distinct class. These distinctions matter because they affect how pixels are translated into the thesis legend.

The source data used in the project contain Philippine-specific categories including Closed Forest, Open Forest, Mangrove Forest, Brush/Shrubs, Grassland, Annual Crop, Perennial Crop, Open/Barren, Built-up, Marshland/Swamp, Fishpond, and Inland Water. These categories provide thematic information that is particularly valuable when adapting a global agreement framework to the Philippine landscape.

The available NAMRIA source was supplied as polygon-based land-cover data and was subsequently merged, rasterised, and incorporated into a project master VRT. During this preparation, a unique-value audit showed that the class values present in the prepared raster did not follow a simple numerical sequence corresponding directly to the published category listing. Consequently, the raster values were inspected explicitly before harmonisation rather than inferred from category order. The resulting value-to-class mapping and the final translation into the MOLCA Philippines legend are documented in Chapter 3.

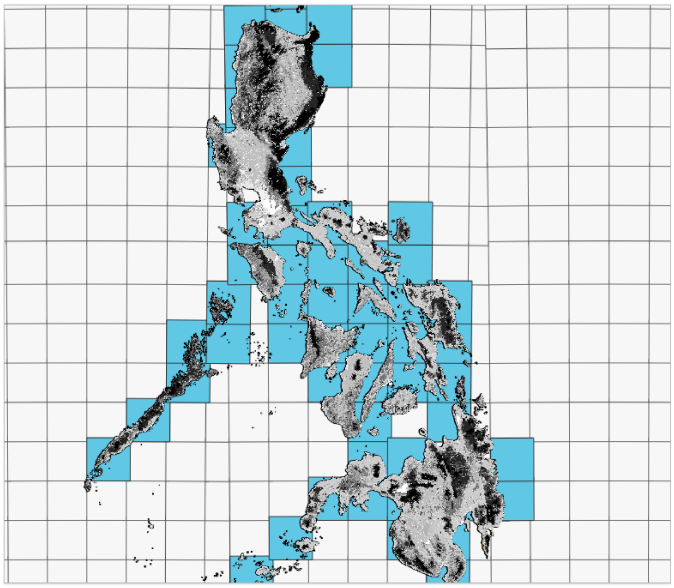


Figure 2.5: NAMRIA land-cover product prepared as a master VRT for the Philippine workflow. The figure illustrates the national coverage used before tile-level clipping and harmonisation.

2.4.2 JAXA High-Resolution Land-Use and Land-Cover Map of Southeast Asia

The second regional product used in the Philippine workflow is the JAXA High-Resolution Land-Use and Land-Cover Map of Southeast Asia. The product version used in the project is 2023SEA_v25.09, representing 2023 and released in September 2025. JAXA reports an Overall Accuracy of 94.59% and a kappa coefficient of 0.94 based on 5,688 test samples [31]. JAXA also notes that this evaluation is preliminary and that the validation samples were not obtained through strict random sampling. These values therefore characterise the reported evaluation of the JAXA product itself and are not accuracy estimates for the MOLCA-derived reference or the deep-learning model developed in this thesis. The product is distributed as GeoTIFF tiles in geographic coordinates (WGS84), with a mesh size of 1/12000 degree, corresponding to approximately 10 m spatial sampling [31].

In the thesis workflow, the JAXA Southeast Asia map was used as a regional land-cover product together with NAMRIA. The two regional products were not introduced to replace the global products. Instead, they were introduced to complement them, especially in pixels where the original global agreement produced AnA. This is important because the thesis remains MOLCA-derived: the objective is still to produce a reliable reference map based on agreement, not to accept a

single product uncritically.

The 15-class JAXA legend contains several categories directly relevant to the Philippine landscape, including Water Bodies, Built-up, Cropland, single- and multi-crop paddy fields, Herbaceous Wetland, Grassland, multiple forest categories, Rubber Tree Plantation, Oil Palm Tree Plantation, Mangrove Forest, and Bare surfaces. The presence of explicit Mangrove and plantation categories makes the product particularly useful for regional adaptation of the reference map.

The JAXA Southeast Asia product is used together with NAMRIA as a regional agreement source rather than as an unconditional replacement for the global products. The exact harmonisation decisions, including the treatment of paddy fields, plantation classes, Mangrove, and other regional categories, are documented in Chapter 3.

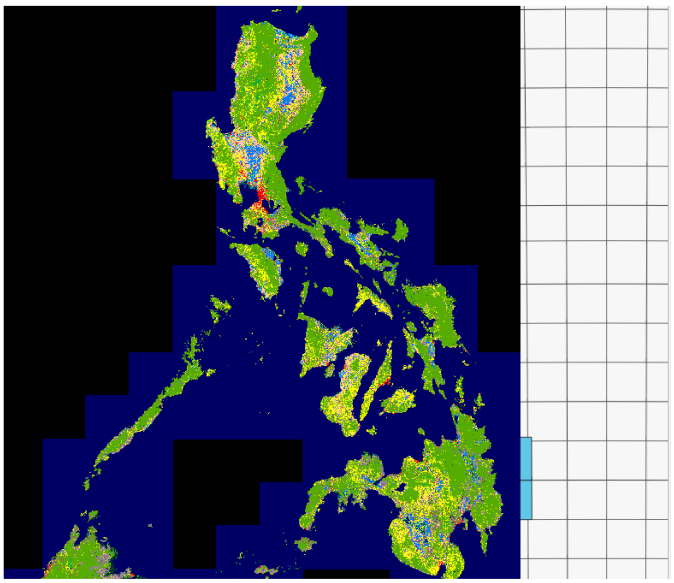


Figure 2.6: JAXA Southeast Asia high-resolution land-use/land-cover product prepared as a master VRT for the Philippine workflow. The product provides regional land-cover information designed for Southeast Asian landscapes and complements the global MOLCA inputs.

Table 2.3: Regional datasets introduced to adapt the MOLCA-derived reference to the Philippine context.

Dataset	**Producer**	**version**	**Role in this thesis**
NAMRIA Land Cover Map	National Mapping and Resource Information Authority, Philippines	2020	National land-cover product used to introduce Philippine-specific class information, including closed/open forest, mangrove forest, brush/shrubs, fishpond, marshland/swamp, and national crop classes.
JAXA HRLULC Southeast Asia	JAXA EORC	2023, released September 2025	Regional 10 m Southeast Asian product used together with NAMRIA to support regional agreement, including Mangrove as a separate class and additional AnA reduction where both regional products agree.

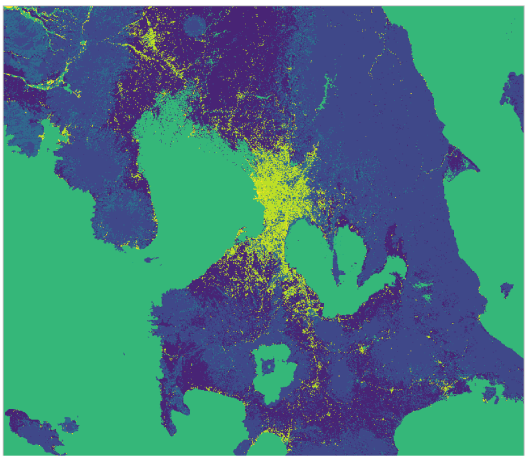
2.5 Comparison of the Dataset Roles

The eight input products do not have the same function. Treating them as equivalent would hide the logic of the workflow. FROM-GLC and GlobeLand30 provide broad land-cover classes across the full study area. The thematic products provide class-specific evidence. NAMRIA and JAXA Southeast Asia provide regional/national information that helps adapt the reference map to the Philippines.

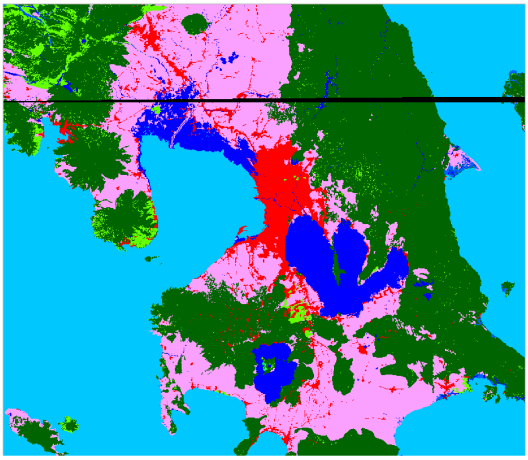
This distinction is summarised in Table 2.4. The table does not describe the full rule hierarchy, because that belongs to Chapter 3. Instead, it explains the conceptual role of each dataset in the thesis.

Table 2.4: Conceptual role of each dataset in the Philippine MOLCA-derived reference design.

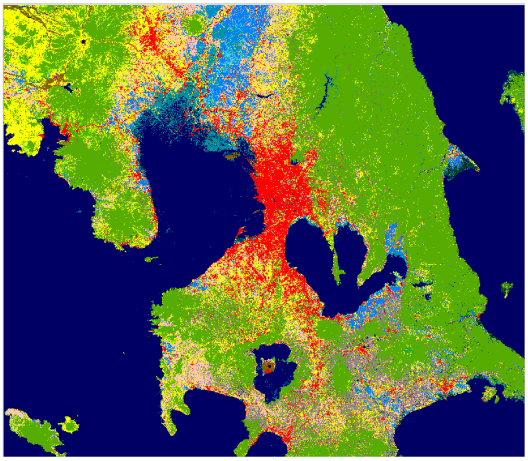
Dataset group	**Datasets**	**Conceptual role**
General global maps	FROM-GLC 2017; GlobeLand30 2020	Provide the principal multi-class land-cover evidence used to form the conservative base agreement.
Built-up evidence	WSF 2019; GHS-BUILT-S	Reinforce the Built-up class where settlement-specific products provide stronger thematic evidence than the general maps alone.
Water/wetland evidence	GSW 2019	Supports water-related classes using a product specifically designed to map surface-water occurrence and history.
Forest evidence	JAXA FNF	Supports Forest using a radar-derived forest/non-forest thematic product.
Regional/national products	NAMRIA 2020; JAXA 2023SEA_v25.09	Improve Philippine specificity, support Mangrove as a separate class, and help reduce AnA where both regional datasets agree.



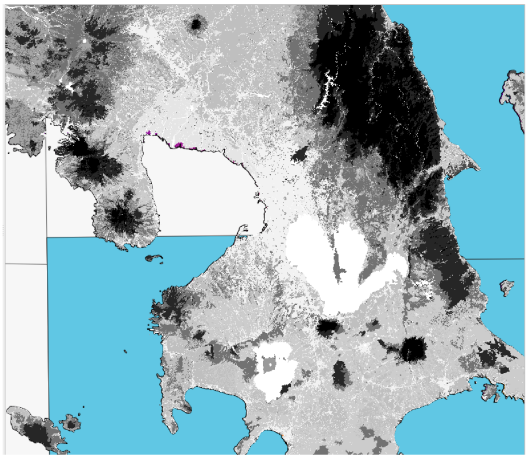
(a) FROM-GLC 2017



(b) GlobeLand30 2020



(c) JAXA Southeast Asia



(d) NAMRIA 2020

Figure 2.7: Example visual comparison over the Manila area for selected general and national/regional products. The purpose of the figure is not to validate the datasets visually, but to illustrate differences in spatial detail, thematic representation, and class structure over the same landscape. Each panel uses source-specific symbology, so colours should not be compared directly between products.

The Manila comparison is useful because it illustrates the central reason for using agreement rather than a single product. Urban, peri-urban, agricultural, and vegetated surfaces are all present within a relatively small area. Some datasets emphasise built-up surfaces, others represent broader land-cover categories, and the national/regional products contain locally relevant classes. The MOLCA-derived strategy uses this diversity as a strength: agreement can identify more reliable

labels, while disagreement is treated explicitly rather than hidden.

2.6 Final Target Legend

All input datasets must be translated into a common legend before they can be compared. Without a shared class system, two datasets may appear to disagree simply because they use different names, class levels, or thematic definitions. For example, one product may distinguish closed forest and open forest, while another only has Forest. One product may include fishpond as a separate class, while another may treat it as Water. Similarly, mangrove may appear as Forest, Wetland, or a separate coastal class depending on the product.

The final target legend used in this thesis contains one unresolved code (NoData/AnA) and nine land-cover classes. Eight land-cover classes are shared with the original MOLCA legend: Built-up, Cropland, Forest, Grassland, Shrubland, Wetland, Water, and Bareland. The original Permanent ice and snow class is not retained for the Philippine application, while Mangrove is introduced as a separate Philippines-specific class [3]. This addition is supported by the explicit mangrove information available in both regional datasets.

Table 2.5: Final target legend used for the Philippine MOLCA-derived reference map.

Class ID	**Class name**	**General interpretation in this thesis**
0	NoData / AnA	Pixels without reliable agreement or outside the valid mapping decision.
1	Built-up	Urban and settlement-related surfaces.
2	Cropland	Annual and perennial agricultural surfaces mapped into the target cropland category.
3	Forest	Forested areas, including closed/open forest and regional forest classes not separated as mangrove.
4	Grassland	Herbaceous grass-dominated land cover.
5	Shrubland	Brush/shrub-dominated land cover.
6	Wetland	Marshland, swamp, and herbaceous wetland-type classes.
7	Water	Inland water, surface water, and fishponds according to the final project decision.
8	Bareland	Open, barren, or sparsely vegetated surfaces.
9	Mangrove	Philippines-specific coastal mangrove class retained separately when supported by regional agreement.

The presence of Class 0 is not a failure of the method. It is part of the logic of a conservative reference map. When input datasets disagree and no accepted evidence rule assigns a class, the pixel is not forced into a label. This protects the training dataset from uncertain labels. In a deep learning context, this is important because noisy labels can affect model training, especially for minority classes. The use of AnA therefore reflects the objective of creating reliable training labels, not a wall-to-wall national land-cover map at the reference-generation stage.

2.7 Sentinel-1 SAR Data

The land-cover reference map described above provides labels. The predictive input for the deep learning model comes from Sentinel-1 SAR imagery. Sentinel-1 is the radar component of the European Copernicus programme. It provides C-band Synthetic Aperture Radar observations and is designed to acquire data during day or night and under all-weather conditions [32, 33]. This property is especially valuable in tropical and maritime regions such as the Philippines, where optical imagery can be strongly affected by cloud cover, haze, and frequent precipitation.

The use of Sentinel-1 in this thesis follows the methodological direction of the Sentinel-1/Swin-Unet land-cover mapping architecture developed by Russo et al. [9]. That work showed how seasonal Sentinel-1 spatio-temporal features can be used with a transformer-based segmentation model for high-resolution land-cover classification. The Philippine thesis adapts this idea to a new study area and combines it with a locally adapted MOLCA-derived reference.

For the Philippine workflow, Sentinel-1 data were selected for the year 2019. The configuration used descending-orbit acquisitions in Interferometric Wide Swath mode and VH polarisation. VH was selected because cross-polarised radar backscatter is sensitive to vegetation structure and land-cover differences, and it has been used in previous Sentinel-1 land-cover mapping work [9]. The detailed preprocessing and feature extraction are not described in this chapter because they belong to Chapter 4. However, the dataset-level description is needed here to explain what remote-sensing observations support the classification.

The final Sentinel-1 representation used in the project is a multi-temporal feature stack. The 2019 acquisition sequence is divided into four fixed temporal groups, and seven feature representations are generated for each group: SI, the Lee-filtered SI, MAX, MIN, MAXMIN, MEAN, and MEDIAN.

$$N_{\text{features}} = N_{\text{groups}} \times N_{\text{features/group}} = 4 \times 7 = 28. \tag{2.1}$$

Equation 2.1 is included only to define the dataset structure. The mathematical details of the seasonal super image, speckle filtering, and spatial feature computation are given in Chapter 4.

Table 2.6: Sentinel-1 dataset configuration used as predictor data in the thesis.

Item	Configuration in this thesis
Mission	Copernicus Sentinel-1.
Sensor type	C-band Synthetic Aperture Radar.
Product family	Sentinel-1 GRD imagery used for land-cover feature generation.
Observation year	2019.
Acquisition mode	Interferometric Wide Swath.
Orbit direction	Descending.
Polarisation	VH.
Temporal organisation	Four temporal groups.
Feature representation	Seven representations per temporal group: SI, Lee-filtered SI, MAX, MIN, MAXMIN, MEAN, and MEDIAN.
Total input channels	28 Sentinel-1 feature layers per patch.
Detailed processing chapter	Chapter 4.

2.8 Coordinate Reference Systems and Spatial Resolution

A land-cover agreement map can only be meaningful if all input rasters are spatially comparable. The input products used in this thesis differ in spatial resolution, native projection, tiling, and class definition. Some products are distributed at 10 m, while others are distributed at 25 m or 30 m. Some are general land-cover maps, while others are binary or thematic layers. Some are global, while others are regional or national. These differences make preprocessing a scientific requirement rather than a purely technical step.

The target spatial resolution of the MOLCA-derived reference map is 10 m. This choice is consistent with the original MOLCA framework and with the use of Sentinel-1 and Sentinel-2 high-resolution Earth Observation products [3, 9]. In the Philippine implementation, FROM-GLC provides the reference grid for alignment. Other rasters are reclassified and aligned to this grid before agreement computation. This prevents false disagreement caused by pixel shifts, different grid origins, or different raster extents.

The detailed alignment method, including clipping, reclassification, GDAL/GRASS processing, and agreement computation, is explained in Chapter 3. Here, it is enough to state the data requirement: all input datasets must be made comparable in class system and spatial grid before they can contribute to a reference label.

Table 2.7: Main spatial-resolution considerations for the datasets used in this thesis.

Spatial-data group	**Relevance to the workflow**
Native or nominal 10 m raster products	FROM-GLC 2017, GHS-BUILT-S E2018, WSF 2019, and JAXA 2023SEA_v25.09 provide data at, or approximately at, the target 10 m mapping scale.
Native 25–30 m raster products	JAXA FNF 2019 (25 m), GlobeLand30 2020 (30 m), and JRC Global Surface Water Yearly History 2019 (30 m) provide useful thematic or land-cover information but require alignment to the 10 m reference grid.
Vector or project-derived sources	The NAMRIA 2020 source was supplied as polygon-based land-cover data and was rasterised during project preparation. Sentinel-1 feature rasters are generated later in the workflow for the selected model patches.
Common project grid	Before agreement computation, all rasterised and reclassified reference layers are aligned to the 10 m FROM-GLC reference grid.

2.9 Chapter Summary

This chapter defined the study area and the datasets used in the thesis. The Philippines was selected as a complex tropical and archipelagic study area where high-resolution land-cover mapping is both scientifically relevant and technically challenging. The Sentinel-2 tiling grid was adopted as the spatial organisation framework because it provides reproducible tile-level units for processing, validation, patch selection, and later Sentinel-1 feature generation.

The datasets were divided into three functional groups. First, FROM-GLC 2017 and GlobeLand30 2020 provide the principal global land-cover maps used to form the base agreement. Second, GHS-BUILT-S E2018, WSF 2019, JRC Global Surface Water Yearly History 2019, and JAXA FNF 2019 provide thematic evidence for Built-up, Water/Wetland, and Forest. Third, NAMRIA 2020 and JAXA 2023SEA_v25.09 provide national and regional land-cover information that supports the Philippine adaptation, particularly the separate Mangrove class and the reduction of remaining AnA where the two additional products agree.

The chapter also introduced the final target legend and the Sentinel-1 dataset configuration. The target legend contains NoData/AnA and nine land-cover classes: Built-up, Cropland, Forest, Grassland, Shrubland, Wetland, Water, Bareland, and Mangrove. Sentinel-1 SAR data from 2019 provide the predictor information for the deep learning model, organised into 28 multi-temporal Sentinel-1 feature layers. The next chapter explains how the datasets described here are harmonised, reclassified, spatially aligned, and combined to generate the MOLCA-derived reference map for the Philippines.

Chapter 3

Generation of the MOLCA Philippines Reference Map

3.1 Purpose of the Chapter

The previous chapter introduced the Philippine study area and described the datasets used in the thesis. This chapter explains how those datasets were transformed into a single MOLCA-derived reference map for the Philippines. The chapter is therefore methodological, but it is also central to the scientific structure of the thesis: the deep learning model described later can only learn from the labels that are provided to it. If the reference map is inconsistent, spatially misaligned, or thematically ambiguous, the model will inherit these weaknesses. For this reason, the construction of the reference map is treated as a major part of the thesis rather than as a simple preprocessing step.

The aim of this chapter is to describe, step by step, how the input land-cover products were harmonised, aligned, reclassified, and combined through agreement logic. The workflow follows the general principle of the Map of Land Cover Agreement (MOLCA), where existing high-resolution land-cover products are combined to retain pixels with stronger label reliability [3]. The Philippine implementation keeps this agreement-based philosophy, but adapts it to the available global, thematic, national, and regional datasets introduced in Chapter 2. In particular, the workflow uses two general land-cover products, four thematic evidence products, and one national product and one regional product to generate a 10 m reference map organised by Sentinel-2 tiles.

This chapter focuses on the reference-map generation workflow. It does not describe the selection of 549 $\times$ 549 patches, because that belongs to Chapter 4. It also does not describe the Sentinel-1 feature extraction pipeline, which is part of Chapter 4. Finally, it does not evaluate the deep learning classification results,

which are discussed in Chapter 6. This separation is important because the thesis contains three related but distinct operations: first, generation of the MOLCA Philippines reference map; second, selection of training and validation areas and preparation of Sentinel-1 features; third, semantic segmentation using deep learning.

3.2 Terminology: AnA and NaN

Before describing the reference-map generation workflow, it is necessary to clarify the terminology used in this chapter. In the methodological discussion, we use the term *AnA* to indicate pixels for which no reliable land-cover label is assigned after the agreement procedure. In the context of this thesis, AnA can be interpreted as *Area not Available* or, more specifically, as an area where the available land-cover products do not provide sufficient agreement for a confident reference label.

In the practical outputs generated by the processing scripts, these unresolved pixels are counted in CSV fields named `NaN_Pixels` and `NaN_Percentage`. In those reports, *NaN* is therefore a software/reporting label for unresolved pixels rather than a separate land-cover class.

For consistency, the thesis uses *AnA* when discussing the reference-map logic and *NaN* only when referring directly to the CSV outputs or software-generated tables. This terminology must also be distinguished from source-specific code 0 and from raster NULL/NoData values encountered during preprocessing. Source code 0 is interpreted according to the individual product, while GRASS/GDAL NULL or NoData represents missing or invalid raster data. After harmonisation and agreement processing, unresolved reference pixels are represented by target code 0 and described conceptually as AnA.

3.3 Conceptual Basis of MOLCA

A land-cover reference map used for training a deep learning model should satisfy two conditions. First, it should contain a sufficient number of labelled pixels for the classes of interest. Second, the labels should be reliable enough that the model learns meaningful class boundaries rather than noise. These two requirements are often in tension. A complete wall-to-wall product provides labels everywhere, but some of those labels may be uncertain. A manually checked reference set can be highly reliable, but it is often too small for deep learning. MOLCA addresses this problem by using agreement among existing land-cover products as a proxy for label reliability.

The basic idea is simple. Suppose that several products describe the same

pixel. If two or more products assign the same class to that pixel, the agreement provides stronger support for that label than a single product alone. If products disagree and no supporting evidence exists, the pixel should not be forced into a class. It can instead be left as NoData or Area not Available (AnA). In this thesis, AnA has an important meaning: it indicates that the available sources did not provide sufficient agreement to support a training label.

The original MOLCA framework was proposed as a way to generate large-scale high-resolution reference data by integrating existing land-cover products [3]. Its contribution is not the creation of a new classifier, but the creation of an agreement-derived reference layer from already available maps. The approach is suitable for deep learning because segmentation models require large numbers of labelled pixels. Rather than manually labelling every pixel, MOLCA retains pixels where the participating products satisfy the defined agreement conditions.

For the Philippine work, the same principle was adapted to the national study area. The method can be summarised as follows:

1. define a common target legend;
2. bring all input products to a common spatial grid;
3. reclassify the products to the common legend;
4. compute a base agreement using general land-cover maps;
5. use thematic products to support specific classes;
6. use national and regional products to improve the local relevance of the reference map;
7. keep unresolved pixels as AnA rather than assigning uncertain labels.

This philosophy is conservative. It does not attempt to produce the most complete possible map at the reference-generation stage. It attempts to produce a map whose labels are sufficiently reliable for downstream training. The distinction matters because a deep learning model trained on incorrect labels may learn systematic errors. In this thesis, class reliability is therefore preferred over full spatial coverage during reference generation.

3.4 From General MOLCA to MOLCA Philippines

The original MOLCA methodology was designed for large geographical regions and used a combination of global, thematic, and regional high-resolution land-

cover products [3]. The Philippine workflow follows the same general logic but modifies the input set and class treatment according to the study area.

The initial Philippine implementation used the six products closest to the original MOLCA design: FROM-GLC 2017, GlobeLand30 2020, WSF 2019, GHS-BUILT-S, GSW 2019, and JAXA FNF. These products provided a strong global foundation. FROM-GLC and GlobeLand30 were used as the two multi-class global land-cover sources. WSF and GHS-BUILT-S supported the built-up class. GSW supported water and wetland-related decisions. FNF supported forest. This initial structure was useful, but it was not enough to fully exploit Philippine-specific information.

The workflow was later extended with two additional products: the national NAMRIA 2020 land-cover map and the regional JAXA High-Resolution Land-Use and Land-Cover Map of Southeast Asia 2023. These products were not added to replace the global products. They were added to improve the local representation of classes that are better described by Philippine national or Southeast Asian mapping systems. This is especially important for mangroves, fishponds, brush/shrub vegetation, and plantation classes. In the final target legend, Mangrove was retained as a separate Class 9.

Table 3.1 summarises the relationship between the original MOLCA idea and the Philippine implementation.

Table 3.1: Conceptual adaptation of the MOLCA framework to the Philippine workflow.

Component	**General MOLCA principle**	**Philippine implementation**
Reference objective	Extract reliable labelled pixels from agreement among existing HRLC products.	Generate a 10 m MOLCA-derived reference map for the Philippines to support deep learning land-cover segmentation.
Spatial organisation	Large-area processing divided into manageable spatial units.	Sentinel-2 tiles used as operational units; the final workflow was applied tile by tile.
General land-cover maps	Multi-class products provide the base land-cover agreement.	FROM-GLC 2017 and GlobeLand30 2020 provide the base agreement.
Thematic products	Class-specific products reinforce selected classes.	WSF and GHS-BUILT-S reinforce Built-up; GSW supports Water/Wetland; FNF supports Forest.
National/regional products	Locally relevant products can improve agreement in specific geographical contexts.	NAMRIA 2020 and JAXA Southeast Asia 2023 support Philippine-specific adaptation, including Mangrove.
Uncertain pixels	Pixels without sufficient agreement are not forced into a class.	Remaining disagreement pixels are coded as NoData/AnA.

3.5 Tile-Based Processing Strategy

The Philippine study area is large and geographically fragmented. Processing all input rasters for the full country in one operation would be inefficient and difficult to validate. The workflow was therefore designed around Sentinel-2 tiles. Each tile is processed independently, using the same sequence of operations. This provides a clean unit for clipping, alignment, reclassification, agreement computation, validation, and later patch extraction.

The tile-based design has four advantages. First, it reduces memory and storage requirements because the workflow handles one spatial unit at a time. Second, it

makes errors easier to diagnose. If a problem appears in one tile, the corresponding input layers and intermediate outputs can be inspected locally. Third, it allows the workflow to be tested on pilot tiles before being applied to the full set. Fourth, it connects naturally to later patch selection and Sentinel-1 feature generation, both of which also use tile and patch identifiers.

The tile used for early pilot testing was 51PUS. This tile was used to test clipping, harmonisation, reclassification, grid alignment, agreement computation, and validation. After the pilot workflow was verified, the procedure was extended to additional tiles and later applied to the selected operational tile set. The project archive contains a 400-point validation structure for 51PUS and additional 800-point logic-validation tables for 51NXG and 51PXP. These files record point coordinates, extracted raster values and validation fields, providing tile-level evidence that the implemented agreement logic was checked at sampled locations.

DATASET	SHAPE	CRS	PIXEL COUNT (VALID)
from	10980 x 10980	EPSG:4326	118,314,427
gl30	10980 x 10980	EPSG:4326	50,515,603
gsw	10980 x 10980	EPSG:4326	70,901,366
fnf	10980 x 10980	EPSG:4326	39,303,473
ghs	10980 x 10980	EPSG:None	37,468,863
wsf	10980 x 10980	EPSG:4326	3,473,540
jaxa	10980 x 10980	EPSG:4326	118,319,959
namria	10980 x 10980	EPSG:4326	50,413,117

Figure 3.1: Example tile-grid evidence used during MOLCA processing. The figure documents the 10980 × 10980 pixel structure of a Sentinel-2 tile at 10 m resolution and supports the tile-based processing design used in the thesis.

3.6 Input Layers Used in the Reference-Map Generation

Chapter 2 described the datasets themselves. This section explains their functional role in the reference-map generation. The workflow uses eight input layers:

1. FROM-GLC 2017;
2. GlobeLand30 2020;
3. GHS-BUILT-S;
4. WSF 2019;
5. GSW 2019;

6. JAXA FNF;

7. NAMRIA 2020;

8. JAXA Southeast Asia HRLULC 2023.

The first two products form the base multi-class agreement. The next four products act as thematic evidence layers. The final two products provide national and regional information for the Philippines. This structure is shown in Figure 3.2.

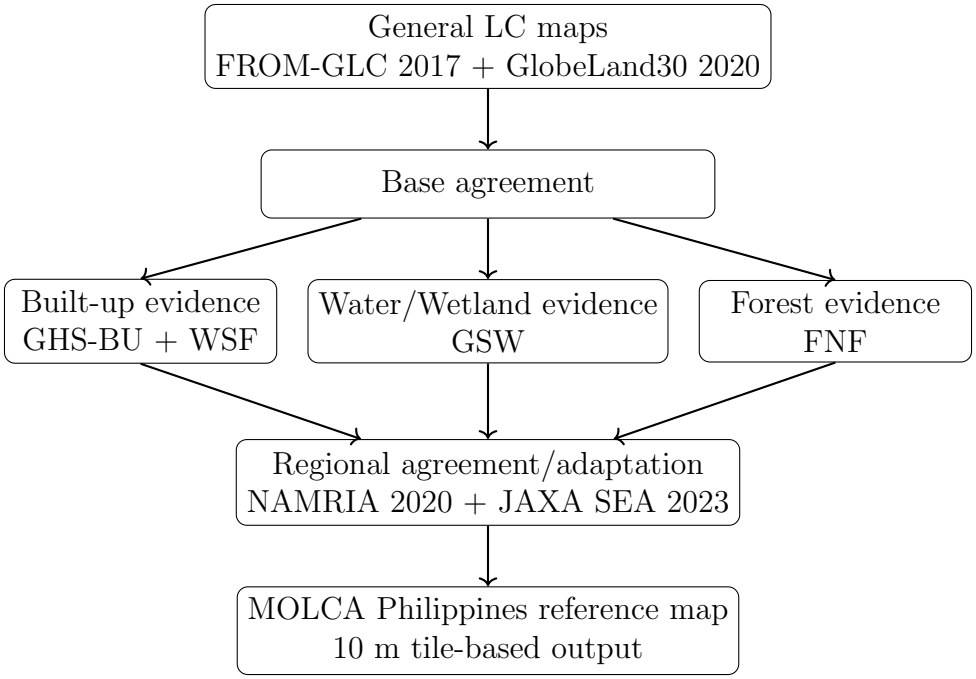


Figure 3.2: Functional grouping of the eight reference sources. The diagram shows the role of each dataset family and does not represent the sequential priority of the agreement hierarchy, which is shown in Figure 3.5.

3.7 Preprocessing Requirements

The input products used in the MOLCA Philippines workflow could not be compared directly in their original form. They were produced by different institutions, distributed in different formats, represented with different class codes, and stored at different spatial resolutions and coordinate reference systems. A pixel-level agreement procedure is only meaningful when all input layers refer to the same spatial unit and to the same target class system. For this reason, preprocessing was a central methodological step rather than a minor technical operation.

The preprocessing stage had two complementary objectives. The first was spatial standardisation. Each source product had to be prepared so that its pixels matched the tile-level reference grid used for the Philippine workflow. This required clipping each dataset to the Sentinel-2 tile extent, aligning the layer to the common 10 m grid, and preserving categorical values through nearest-neighbour logic during resampling. The second objective was semantic standardisation. Each original class code had to be translated into the MOLCA Philippines legend before any agreement operation could be applied.

The practical workflow therefore included the following operations:

1. preparation of master raster or VRT files for each source dataset;
2. clipping of each source layer to the Sentinel-2 tile extent;
3. harmonisation and reclassification of the source legend into the target class IDs;
4. spatial alignment of the reclassified rasters to the common 10 m reference grid using categorical nearest-neighbour resampling;
5. visual inspection of the aligned and reclassified layers;
6. numerical checking through unique-value reports and validation tables;
7. staged agreement computation after the inputs had passed the spatial and thematic checks.

The eight comparison images used in this thesis document this preprocessing stage. They do not represent raw input products. Instead, they show layers that were clipped to the selected test area, aligned to the 10 m grid, and harmonised/reclassified to the MOLCA Philippines class system. This distinction is important because the figures demonstrate that the input products had already been transformed into comparable tile-level layers before the agreement rules were applied.

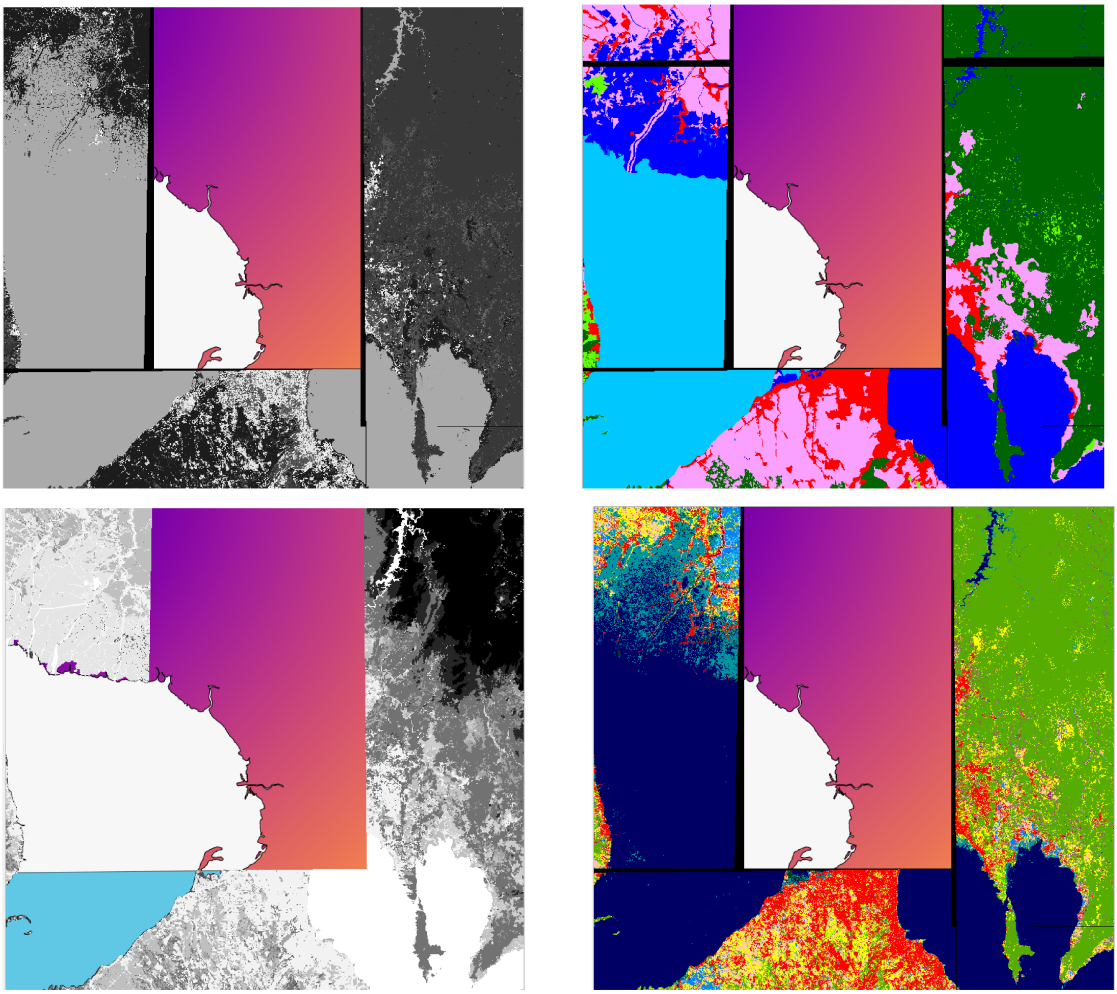


Figure 3.3: Examples of clipped, 10 m aligned, and harmonised/reclassified layers over the four-tile test area. The figure compares the two general global products with the national NAMRIA product and regional JAXA Southeast Asia product used in the Philippine adaptation.

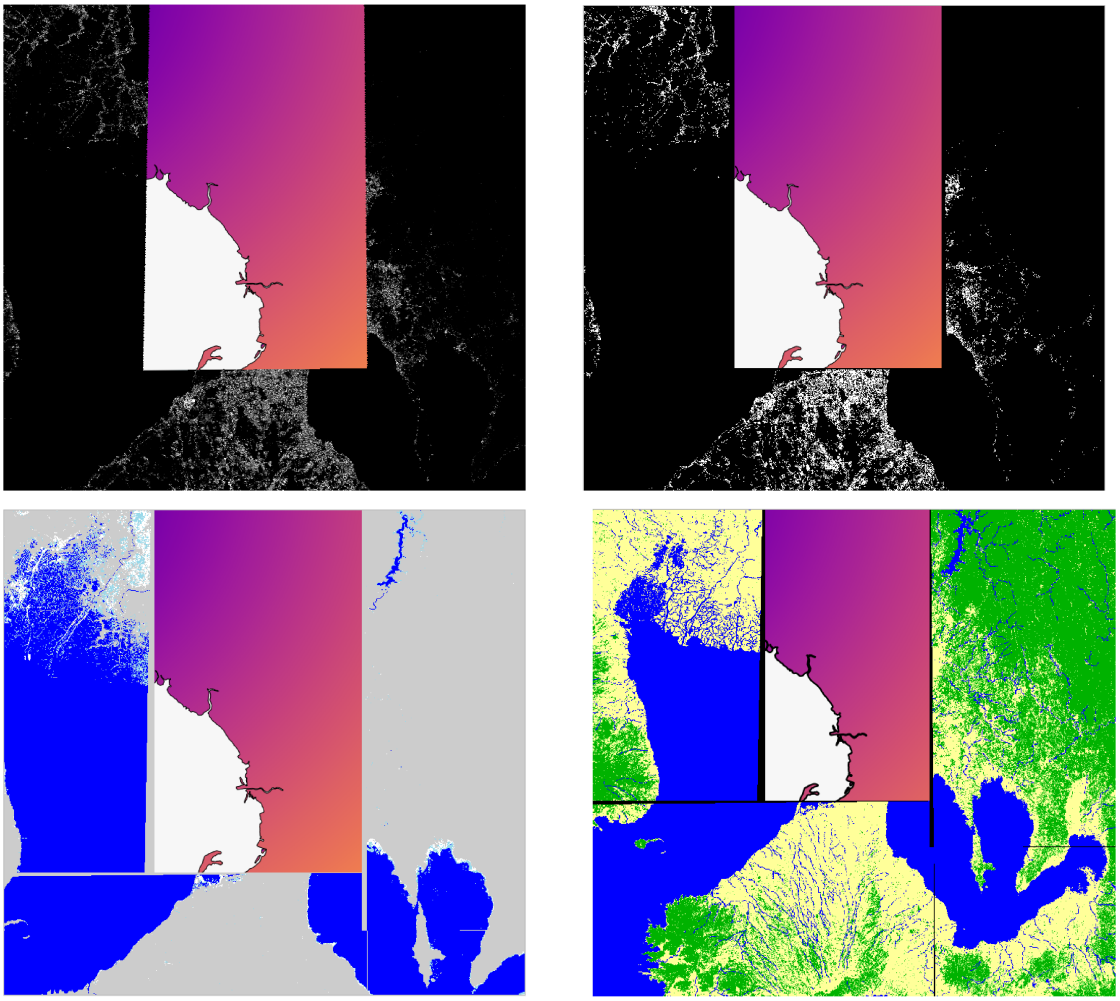


Figure 3.4: Examples of clipped, 10 m aligned, and harmonised/reclassified thematic evidence layers. GHS-BUILT-S and WSF support Built-up, GSW supports the project Water/Wetland mapping, and FNF supports Forest decisions.

3.8 Common Legend and Reclassification Rules

A common legend is required before any agreement rule can be applied. The eight input products used in this thesis do not share the same class system. Some products are complete multi-class land-cover maps, while others are thematic evidence layers designed only for one class group, such as settlement, water, or forest. In addition, the national NAMRIA product and regional JAXA Southeast Asia product, contain locally relevant classes that are not represented in exactly the same way in the global products.

For this reason, all products were harmonised into a common MOLCA Philippines legend before agreement computation. The final legend contains one unresolved class and nine land-cover classes. The unresolved class is coded as 0 and corresponds to NoData/AnA. The land-cover classes are Built-up, Cropland, Forest, Grassland, Shrubland, Wetland, Water, Bareland, and Mangrove.

Table 3.2: Final target legend used during MOLCA Philippines reference-map generation.

Code	**Class**	**Interpretation in the reference map**
0	NoData / AnA	No reliable agreement or invalid/unassigned pixel.
1	Built-up	Settlement, urban, and other built surfaces.
2	Cropland	Annual and perennial agricultural surfaces.
3	Forest	Forested areas except Mangrove when separated by regional agreement.
4	Grassland	Herbaceous grass-dominated areas.
5	Shrubland	Brush/shrub-dominated areas.
6	Wetland	Marshland, swamp, and herbaceous wetland.
7	Water	Inland water, surface water, and fishponds according to the project rule.
8	Bareland	Open, barren, and sparsely vegetated surfaces.
9	Mangrove	Coastal mangrove class introduced for the Philippine adaptation.

The reclassification was not identical for all products. FROM-GLC and GlobeLand30 were mapped as general multi-class land-cover products. WSF and GHS-BUILT-S were used only as Built-up evidence. GSW was used as Water/Wetland evidence. FNF was used as Forest evidence and, where applicable, Water evidence. NAMRIA and JAXA Southeast Asia were used as national/regional agreement sources after their own class systems were translated into the same target legend.

Table 3.3: Reclassification rules for the six original MOLCA-style input datasets used in the Philippine workflow (Part 1).

Dataset	**Original value**	**MOLCA ID**	**Target class**	**Role or note**
FROM-GLC	0	0	NoData	No valid class.
FROM-GLC	10	2	Cropland	General LC class.
FROM-GLC	20	3	Forest	General LC class.
FROM-GLC	30	4	Grassland	General LC class.
FROM-GLC	40	5	Shrubland	General LC class.
FROM-GLC	50	6	Wetland	General LC class.
FROM-GLC	60	7	Water	General LC class.
FROM-GLC	80	8	Bareland	General LC class.
FROM-GLC	90	1	Built-up	General LC class.
GlobeLand30	10	2	Cropland	Cultivated land.
GlobeLand30	20	3	Forest	Forest.
GlobeLand30	30	4	Grassland	Grassland.
GlobeLand30	40	5	Shrubland	Shrubland.
GlobeLand30	50	6	Wetland	Wetland.
GlobeLand30	60	7	Water	Water bodies.
GlobeLand30	70	0	NoData / AnA	Tundra, not retained in the Philippine target legend.
GlobeLand30	80	1	Built-up	Artificial surfaces.
GlobeLand30	90	8	Bareland	Bare land.
GlobeLand30	100	0	NoData / AnA	Permanent snow and ice, not retained.
GlobeLand30	255	0	NoData / AnA	Background, ocean, or NoData.

Table 3.4: Reclassification rules for the six original MOLCA-style input datasets used in the Philippine workflow (Part 2).

Dataset	Original value	MOLCA ID	Target class	Role or note
WSF	255	1	Built-up	Built-up evidence.
WSF	0	0	NoData / AnA	Non-built-up; not evidence for other classes.
WSF	1	0	NoData / AnA	Treated as NoData/noise according to project clarification.
GHS-BU	1–100	1	Built-up	Any built-up fraction used as Built-up evidence.
GHS-BU	0	0	NoData / AnA	No built-up evidence.
GHS-BU	NoData	0	NoData / AnA	Outside coverage.
GSW	3	7	Water	Permanent water / strong water evidence.
GSW	2	6	Wetland	Seasonal or intermittent water.
GSW	1	0	NoData / AnA	No water evidence.
GSW	0	0	NoData / AnA	NoData.
FNF	1	3	Forest	Forest evidence.
FNF	2	0	NoData / no evidence	Non-forest; not assigned to another target class.
FNF	3	7	Water	Source Water class retained during harmonisation; not used by the later Forest-fill stage.
FNF	0 / NoData	0	NoData / no evidence	No valid source information.

3.8.1 NAMRIA Raster-Value Check and Harmonisation

The NAMRIA dataset required particular attention because the class categories described in the source documentation did not directly correspond to the raster

values obtained after rasterisation and VRT preparation. The dataset was originally provided as vector polygon data, with land-cover classes such as Closed Forest, Open Forest, Mangrove Forest, Brush/Shrubs, Grassland, Perennial Crop, Annual Crop, Open/Barren, Built-up, Marshland/Swamp, Fishpond, and Inland Water. After merging the polygon data, rasterising it, and building the master VRT, the unique-value audit showed values including 1, 3, 4, 7, 10, 13, 14, 15, 16, 17, 19, 20, and 21. The project harmonisation evidence assigns target classes to the documented values listed in Table 3.5; source value 3 is not assigned a target land-cover class in that final mapping and therefore does not provide class evidence.

This mismatch meant that the published class order could not be used blindly as the raster reclassification rule. Instead, the actual raster values had to be inspected and then mapped manually to the MOLCA Philippines legend. This step was essential because a wrong interpretation of the NAMRIA raster values would propagate directly into the regional agreement stage and could create systematic reference-label errors.

The final NAMRIA harmonisation table used in this thesis maps the explicitly identified source values for Closed Forest and Open Forest to Forest, Mangrove Forest to Mangrove, Brush/Shrubs to Shrubland, Grassland to Grassland, Annual Crop and Perennial Crop to Cropland, Marshland/Swamp to Wetland, Fishpond and Inland Water to Water, Built-up to Built-up, and Open/Barren to Bareland.

Table 3.5: NAMRIA 2020 harmonisation used in the MOLCA Philippines workflow.

NAMRIA value	NAMRIA class	MOLCA ID	Target class / note
1	Closed Forest	3	Forest.
4	Open Forest	3	Forest.
7	Mangrove Forest	9	Mangrove, retained as a new separate class.
10	Brush/Shrubs	5	Shrubland.
13	Open / Barren	8	Bareland.
14	Grassland	4	Grassland.
15	Marshland / Swamp	6	Wetland.
16	Annual Crop	2	Cropland.
17	Perennial Crop	2	Cropland.
19	Fishpond	7	Water, according to the project decision.
20	Built-up	1	Built-up.
21	Inland Water	7	Water.

3.8.2 JAXA Southeast Asia Harmonisation

The JAXA Southeast Asia product was also harmonised into the MOLCA Philippines legend before it was used in the regional agreement stage. The product used in this thesis is the JAXA High-Resolution Land-Use and Land-Cover Map of Southeast Asia in 2023, version 2023SEA_v25.09, released in September 2025. Its classification scheme contains 15 categories, including Water Bodies, Built-up, Solar Panel, Cropland, paddy fields, Herbaceous Wetland, Grassland, forest classes, Rubber Tree Plantation, Oil Palm Tree Plantation, Mangrove Forest, and Bare.

The JAXA product was valuable because it is a regional high-resolution product designed for Southeast Asia. Its legend includes classes that are particularly relevant to the Philippines, especially Mangrove Forest, paddy fields, plantations, and herbaceous wetlands. However, these classes had to be translated into the final MOLCA Philippines legend. In particular, Rubber Tree Plantation and Oil Palm Tree Plantation were mapped to Forest according to the supervisor-approved project decision, while Mangrove Forest was retained as the separate Mangrove class.

Table 3.6: JAXA Southeast Asia harmonisation used in the MOLCA Philippines workflow.

JAXA value	**JAXA class**	**MOLCA ID**	**Target class / note**
0	No Data	0	NoData / AnA.
1	Water Bodies	7	Water.
2	Built-up	1	Built-up.
3	Solar Panel	1	Built-up.
4	Cropland	2	Cropland.
5	Single-crop Paddy Field	2	Cropland.
6	Multi-crop Paddy Field	2	Cropland.
7	Herbaceous Wetland	6	Wetland.
8	Grassland	4	Grassland.
9	Deciduous Broad-leaved Forest	3	Forest.
10	Evergreen Broad-leaved Forest	3	Forest.
11	Evergreen Needle-leaved Forest	3	Forest.
12	Rubber Tree Plantation	3	Forest according to the project decision.
13	Oil Palm Tree Plantation	3	Forest according to the project decision.
14	Mangrove Forest	9	Mangrove, retained as a new separate class.
15	Bare	8	Bareland.

3.9 Mathematical Notation for the Agreement Map

Let Ω_t be the set of valid pixels in a Sentinel-2 tile t. For every pixel $p \in \Omega_t$, each harmonised input raster provides either a target class code or NoData. We denote the harmonised class values as:

$$F(p) = \text{FROM-GLC class at pixel } p, \tag{3.1}$$

$$G(p) = \text{GlobeLand30 class at pixel } p, \tag{3.2}$$

$$B_{GHS}(p), B_{WSF}(p), W_{GSW}(p), R_{FNF}(p), N(p), J(p), \tag{3.3}$$

where B_{GHS} and B_{WSF} are the built-up evidence layers, W_{GSW} is the water/wetland evidence layer, R_{FNF} is the forest evidence layer, N is the harmonised NAMRIA layer, and J is the harmonised JAXA Southeast Asia layer.

The output MOLCA Philippines class for pixel p is denoted by:

$$M(p) \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}, \tag{3.4}$$

where 0 represents NoData/AnA.

The base agreement map M_0 is computed from the two general land-cover maps:

$$M_0(p) = \begin{cases} F(p), & \text{if } F(p) = G(p) \text{ and } F(p) \neq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3.5}$$

Equation 3.5 is intentionally strict. A pixel is labelled only when the two general maps agree on the same valid class. If they disagree, the pixel is left as AnA until additional thematic or regional evidence is considered.

3.10 Agreement Hierarchy

The agreement hierarchy is the rule sequence that converts the harmonised inputs into the final reference map. In the project outputs, the staged reduction reports use the following stages:

1. Base Agreement;
2. Built-Up;
3. Water/Wetland;
4. Forest;
5. Mangrove;
6. Final MOLCA.

The archived CSV reports use this exact six-stage structure for the example tiles 51PTR, 50PRB, 51PUR, and 51PUS, from `0_Base_Agreement` through `5_Final_MOLCA`. Each stage records classified-pixel counts together with the remaining `NaN_Pixels` and `NaN_Percentage`.

3.10.1 Stage 0: Base agreement

The first stage retains pixels where FROM-GLC and GlobeLand30 agree after reclassification to the common target legend. This produces a conservative initial map. The advantage is high reliability; the disadvantage is that many pixels remain AnA. This is expected, especially in heterogeneous landscapes and coastal areas.

3.10.2 Stage 1: Built-up evidence

The second stage uses built-up evidence from GHS-BUILT-S and WSF. The reason for a dedicated built-up stage is that settlement products are designed specifically to detect human settlements or built surfaces. In mixed Philippine landscapes, urban areas may be small, coastal, linear, or fragmented. A general land-cover product may miss some of these areas or classify them inconsistently. Built-up thematic evidence helps recover such pixels.

Let M_1 be the map after built-up evidence. The rule can be written conceptually as:

$$M_1(p) = \begin{cases} 1, & \text{if } M_0(p) = 0 \text{ and built-up evidence is present,} \\ M_0(p), & \text{otherwise.} \end{cases} \tag{3.6}$$

Here, "built-up evidence is present" means that either harmonised GHS-BUILT-S or WSF supports the Built-up class. The rule is applied only where the previous-stage output is unresolved, so labels already assigned by the base agreement are preserved.

3.10.3 Stage 2: Water and wetland evidence

The third stage uses the harmonised GSW Yearly History layer to support Water and Wetland decisions. This stage is especially important in the Philippines because water occupies a large portion of tile extents, and many tiles include coastal, inland, and island environments. A wrong treatment of water can strongly affect class distributions, patch selection, and later model training.

Let M_2 be the output after water/wetland evidence:

$$M_2(p) = \begin{cases} c_w, & \text{if } M_1(p) = 0 \text{ and GSW assigns } c_w, \\ M_1(p), & \text{otherwise,} \end{cases} \tag{3.7}$$

where c_w is the target class produced by the project harmonisation rule: permanent water is mapped to Water (Class 7), while seasonal water is mapped to

Wetland (Class 6). This stage fills only pixels that remain unresolved after the Built-up stage, so earlier labels are preserved.

3.10.4 Stage 3: Forest evidence

The fourth stage uses the 25 m JAXA FNF product as Forest evidence. The source product distinguishes NoData, Forest, Non-forest, and Water; in this stage, only its Forest evidence is used to fill unresolved pixels. It therefore cannot replace a full multi-class land-cover product, but it is useful when the target class is Forest, and the previous stages have not assigned a label. The forest evidence stage can be represented as:

$$M_3(p) = \begin{cases} 3, & \text{if } M_2(p) = 0 \text{ and FNF supports Forest,} \\ M_2(p), & \text{otherwise.} \end{cases} \tag{3.8}$$

This stage reduces AnA in forested areas while preserving all labels assigned by earlier stages. Forest is a broad class, and the subsequent national/regional steps separately handle Mangrove, where the two additional products provide matching evidence.

3.10.5 Stage 4: Mangrove

Mangrove is a Philippines-specific extension of the target legend. In the original MOLCA-compatible legend, mangroves would usually be absorbed into Forest or Wetland depending on the source product. For this thesis, Mangrove was retained as Class 9 because both NAMRIA and JAXA Southeast Asia provide explicit mangrove information and because mangroves are important coastal ecosystems in the Philippines.

The Mangrove stage is conceptually written as:

$$M_4(p) = \begin{cases} 9, & \text{if } M_3(p) = 0,\ N(p) = J(p) = 9, \\ M_3(p), & \text{otherwise.} \end{cases} \tag{3.9}$$

The supervisor-approved decision was to keep Mangrove as a separate class where NAMRIA and JAXA Southeast Asia agree on Mangrove at a pixel that remains unresolved after the preceding stages. This is one of the most important differences between a generic global MOLCA implementation and the Philippine adaptation.

3.10.6 Stage 5: Final MOLCA

The final stage uses agreement between the national NAMRIA product and the regional JAXA Southeast Asia product to resolve remaining AnA pixels where both support the same valid target class. This stage does not mean that a single regional product is accepted everywhere. The key idea is that regional agreement is used to reduce the remaining area of no agreement while preserving the conservative nature of the reference map.

Let M_5 be the final MOLCA Philippines map:

$$M_5(p) = \begin{cases} N(p), & \text{if } M_4(p) = 0,\ N(p) = J(p),\ N(p) \neq 0, \\ M_4(p), & \text{otherwise.} \end{cases} \tag{3.10}$$

Equation 3.10 expresses the national–regional agreement principle in a compact form: if the previous stages leave a pixel unresolved, and NAMRIA and JAXA agree on a valid target class, that class is assigned. Otherwise, the previous value is retained. This keeps the national/regional products as agreement-based support rather than as unconditional replacements for the global workflow.

The final output map for tile t is:

$$M_t(p) = M_5(p), \quad p \in \Omega_t. \tag{3.11}$$

3.11 Workflow Diagram

Figure 3.5 summarises the complete Chapter 3 workflow and provides a compact view of the processing sequence before the detailed stage descriptions.

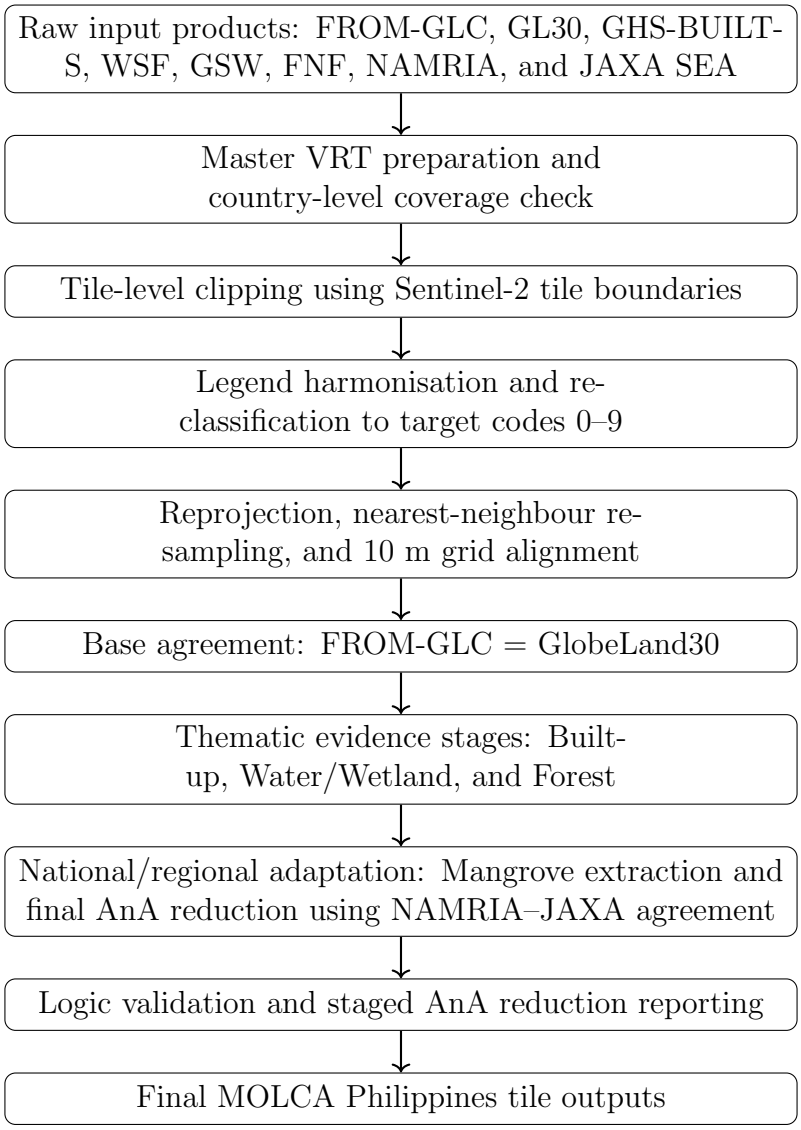


Figure 3.5: Complete MOLCA Philippines reference-map generation workflow. The diagram separates raw datasets, spatial preprocessing, semantic harmonisation, agreement computation, national/regional adaptation, quality checking, and final output production.

3.12 Quantitative Evidence from AnA Reduction Reports

A major advantage of the staged workflow is that each processing stage can be checked quantitatively. In the conceptual description of the thesis, unresolved

pixels are called AnA. In the project CSV files, the same unresolved pixels are reported using the software term NaN, with columns such as `NaN_Pixels` and `NaN_Percentage`. For this reason, the tables in this section keep the CSV terminology when reporting numerical outputs, but the interpretation remains AnA reduction.

The AnA reduction reports count the number of classified pixels and the number of remaining unresolved pixels after each processing stage. These reports make the effect of each evidence layer visible. They also show that the final MOLCA Philippines map is not created in one step, but through a sequence of progressively applied agreement and evidence rules.

Table 3.7 reports the staged reduction for four example tiles: 51PTR, 50PRB, 51PUR, and 51PUS. These values are taken from the CSV files in the Google Drive evidence folder. For 51PTR, the report shows that the base agreement classified 33,226,538 pixels and left 87,333,862 NaN pixels (unresolved AnA pixels); by the final MOLCA stage, classified pixels increased to 113,986,830 and NaN pixels decreased to 6,573,570. Similar staged reductions are recorded for 50PRB, 51PUR, and 51PUS.

Table 3.7: Staged reduction of unresolved AnA pixels for four example tiles. The project CSV files report these unresolved pixels as `NaN_Pixels` and `NaN_Percentage`.

Tile	Stage	Classified pixels	AnA pixels	NaN percentage
51PTR	Base Agreement	33,226,538	87,333,862	72.44
51PTR	Built-up	58,619,028	61,941,372	51.38
51PTR	Water/Wetland	111,186,344	9,374,056	7.78
51PTR	Forest	112,343,817	8,216,583	6.82
51PTR	Mangrove	112,379,338	8,181,062	6.79
51PTR	Final MOLCA	113,986,830	6,573,570	5.45
50PRB	Base Agreement	40,684,489	79,875,911	66.25
50PRB	Built-up	65,282,695	55,277,705	45.85
50PRB	Water/Wetland	102,979,043	17,581,357	14.58
50PRB	Forest	105,831,793	14,728,607	12.22
50PRB	Mangrove	105,864,296	14,696,104	12.19
50PRB	Final MOLCA	110,319,026	10,241,374	8.49
51PUR	Base Agreement	47,142,590	73,417,810	60.90
51PUR	Built-up	68,367,455	52,192,945	43.29
51PUR	Water/Wetland	111,580,220	8,980,180	7.45
51PUR	Forest	112,430,255	8,130,145	6.74
51PUR	Mangrove	112,708,153	7,852,247	6.51
51PUR	Final MOLCA	113,884,894	6,675,506	5.54
51PUS	Base Agreement	47,703,306	72,857,094	60.43
51PUS	Built-up	67,705,640	52,854,760	43.84
51PUS	Water/Wetland	112,156,832	8,403,568	6.97
51PUS	Forest	113,454,886	7,105,514	5.89
51PUS	Mangrove	113,996,491	6,563,909	5.44
51PUS	Final MOLCA	114,502,793	6,057,607	5.02

The table shows a consistent pattern. The base agreement is strict and therefore leaves a large fraction of the tile unresolved. Built-up evidence reduces NaN pixels, but the largest reduction occurs after Water/Wetland evidence. Forest and Mangrove further reduce unresolved areas, and the final MOLCA stage produces the lowest NaN percentage. In the four example tiles, final NaN percentages range from 5.02% to 8.49%. This demonstrates that the final workflow preserves a conservative agreement structure while substantially improving coverage through thematic and regional evidence.

Table 3.8: Overall AnA reduction from base agreement to final MOLCA for the four example tiles.

Tile	Base NaN %	Final NaN %	Reduction	Final classified pixels
51PTR	72.44	5.45	66.99 percentage points	113,986,830
50PRB	66.25	8.49	57.76 percentage points	110,319,026
51PUR	60.90	5.54	55.36 percentage points	113,884,894
51PUS	60.43	5.02	55.41 percentage points	114,502,793

3.13 Staged AnA Reduction Figure

The staged CSV reports are also summarised graphically in Figure 3.6. The plot uses the reported `NaN_Percentage` values for the same four example tiles and shows the contribution of each successive stage to reducing unresolved pixels.

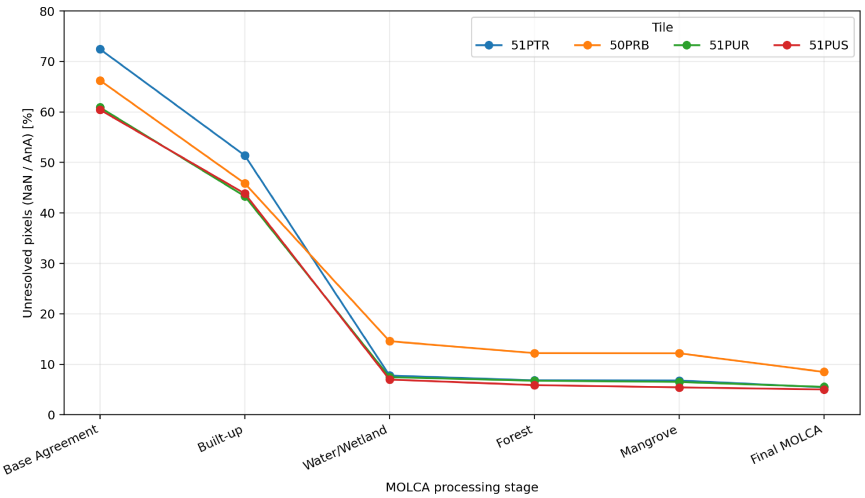


Figure 3.6: AnA percentage after each MOLCA processing stage for four example tiles. The staged pattern shows how base agreement, thematic evidence, Mangrove extraction, and final regional agreement progressively reduce unresolved pixels.

The figure was generated from the same staged values reported in Table 3.7, ensuring consistency between the tabular and graphical summaries.

3.14 Visual Inspection of Final MOLCA Tile Outputs

Quantitative reports are not sufficient on their own. Land-cover reference maps also require spatial inspection because errors can appear as unrealistic shapes, edge artefacts, incorrect coastline patterns, or isolated noisy pixels. For this reason, selected final MOLCA tile outputs were exported as quicklook images and inspected visually.

The quicklook images used in this section correspond to the standard final MOLCA agreement outputs before the later Shrubland rescue experiment. They should not be confused with the separate Shrubland rescue outputs stored in the later project folder. The standard final MOLCA quicklooks belong to the folder `Final_Molca_output_with_csv_and_reports`, while the later Shrubland rescue quicklooks belong to `shrubland_Final_Molca_output_with_csv_and_reports`. The Shrubland rescue material is discussed later in the thesis together with the deep learning diagnostic results.

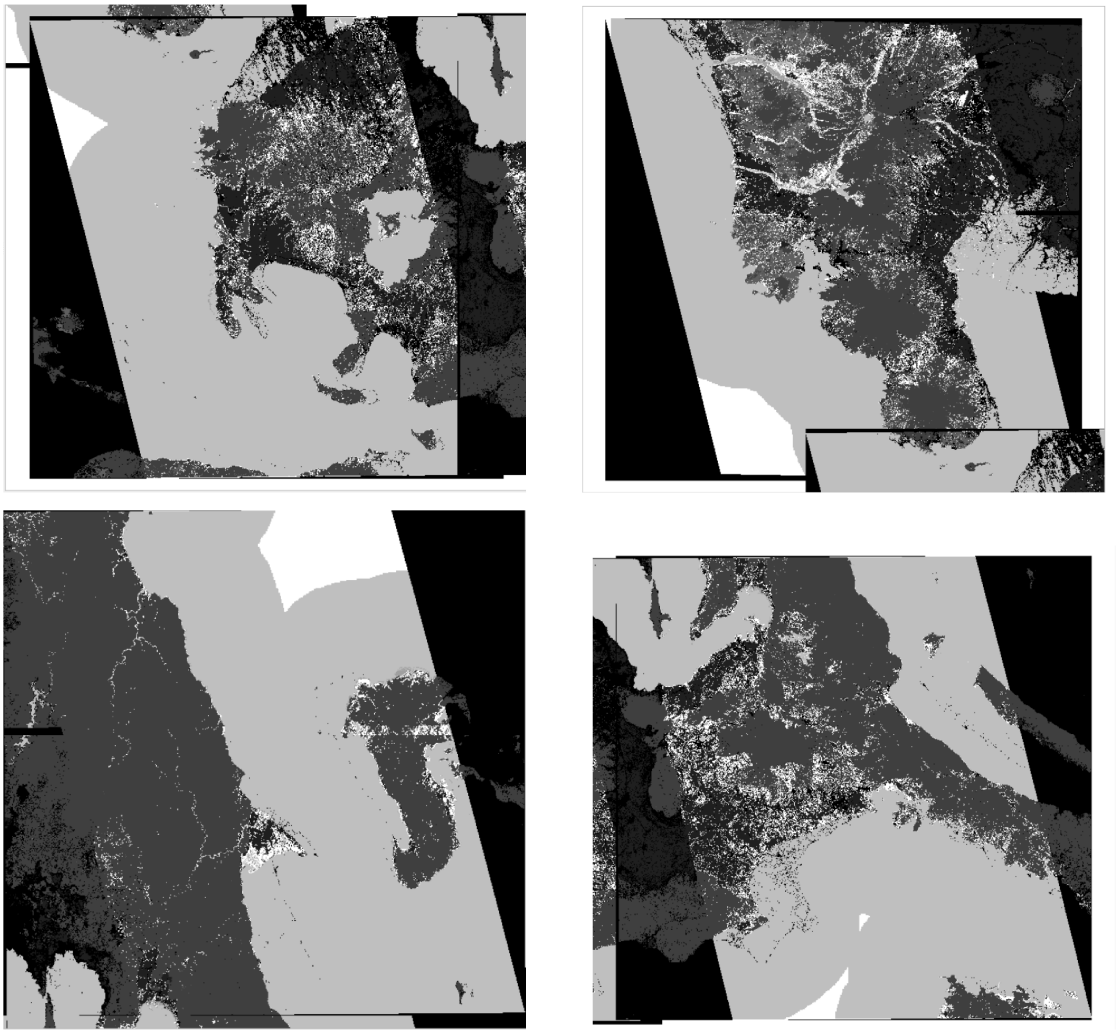


Figure 3.7: Example standard final MOLCA tile outputs before the later Shrubland rescue experiment. These quicklook maps complement the numerical AnA reduction reports by showing spatial patterns in the final agreement outputs.

These quicklooks should be interpreted as visual documentation of the final agreement outputs, not as formal accuracy assessment. Their role is to show that the tile products were inspected spatially and that the final MOLCA layers could be communicated clearly before being used in later patch-selection and deep learning stages.

3.15 Logic Validation and Quality Assurance

The validation performed during the MOLCA generation stage was not a conventional thematic accuracy assessment of the final land-cover product against an independent ground-truth survey. Instead, the goal was to verify that the software correctly implemented the intended rule hierarchy. This distinction is important. At this stage, the question was:

> Does the final output label follow the implemented MOLCA agreement rules for the input values at each sampled point?

This is different from asking whether the land-cover class is correct in the real world according to manual interpretation. The latter belongs to a full thematic validation exercise. The former is a software and logic validation exercise, and it is exactly what was needed before scaling the workflow to additional tiles.

The validation spreadsheets contain sampled points, coordinates, extracted values from the final MOLCA map and input layers, Google Map links, and a `Validation Check` field. For 51NXG, 9 of 800 sampled points are marked INVALID, corresponding to the spreadsheet validation value of 98.875%. For 51PXP, 7 of 800 points are marked INVALID, corresponding to 99.125%. Both files also contain a separate summary field labelled `Agreement` with a value of 100. This `Agreement` field is distinct from the VALID/INVALID validation percentage and should not be interpreted as a 100% thematic accuracy result.

Table 3.9 summarises the validation evidence that can be stated from the available spreadsheets.

Table 3.9: Logic-validation evidence available from the project spreadsheets.

Tile	Validation file	Evidence available in the spreadsheet
51PUS	`validation_51PUS_400 (Editable).xlsx`	Contains 400-point validation structure with coordinates, sample IDs, class layers, reference fields, notes, and Google links.
51NXG	`51NXG_Validation.xlsx`	800 sampled points; 9 marked INVALID (98.875% validation). A separate spreadsheet summary field labelled Agreement has value 100.
51PXP	`51PXP_Validation.xlsx`	800 sampled points; 7 marked INVALID (99.125% validation). A separate spreadsheet summary field labelled Agreement has value 100.

The validation workflow supports three conclusions. First, the output was checked at sampled locations rather than accepted blindly. Second, the check used the actual input values from the harmonised layers, which makes it a direct test of the implemented rule logic. Third, the validation was performed on more than one tile, reducing the risk that the workflow worked only for the pilot tile.

3.16 Role of AnA in the Final Reference Map

AnA is not a missing-data error in this workflow. It is a deliberate output category. When the available products do not provide sufficient agreement, assigning a class would reduce the reliability of the reference map. The reference map is designed for training and evaluating deep learning segmentation. In that context, uncertain labels are dangerous because the model may learn from them as if they were correct. Keeping uncertain pixels as AnA is therefore a form of label-quality control.

The staged reports show that the initial base agreement leaves a large number of pixels unresolved. This is expected because FROM-GLC and GlobeLand30 have different resolution, class definitions, source imagery, and classification methods. The thematic and national/regional stages reduce AnA, but the final map still keeps a small percentage unresolved. In the four archived example reports, the final `NaN_Percentage` values are 5.45% for 51PTR, 8.49% for 50PRB, 5.54% for 51PUR, and 5.02% for 51PUS.

This residual AnA is useful. It identifies locations where the input maps remain uncertain or inconsistent. During later patch selection, AnA pixels can be excluded, masked, or treated carefully depending on the training strategy. This helps prevent the model from being trained on weak labels.

3.17 Automation and Reproducibility

The workflow was designed to be repeatable across tiles. Reproducibility is important because the reference map is not produced manually tile by tile. The same operations must be applied consistently across the study area. The automation logic can be described as a tile-level function:

$$\mathcal{P}(t) \rightarrow M_t, \tag{3.12}$$

where $\mathcal{P}$ is the processing pipeline and M_t is the final MOLCA map for tile t. The pipeline performs the following operations:

1. read the tile identifier and spatial boundary;

2. clip each master dataset to the tile;
3. align all rasters to the reference grid;
4. apply source-specific reclassification rules;
5. compute base agreement;
6. apply thematic evidence stages;
7. apply regional Mangrove and final AnA reduction logic;
8. write the final tile GeoTIFF;
9. export summary statistics and validation support files.

A compact pseudocode version is shown below.

```
for tile_id in selected_tiles:

    # 1. Prepare tile workspace
    tile_boundary = load_sentinel2_tile(tile_id)

    # 2. Clip and align input datasets
    fromglc = clip_reclassify_align(FROM_GLC, tile_boundary)
    gl30    = clip_reclassify_align(GL30, tile_boundary)
    ghsbu   = clip_reclassify_align(GHS_BUILT_S, tile_boundary)
    wsf     = clip_reclassify_align(WSF, tile_boundary)
    gsw     = clip_reclassify_align(GSW, tile_boundary)
    fnf     = clip_reclassify_align(FNF, tile_boundary)
    namria  = clip_reclassify_align(NAMRIA, tile_boundary)
    jaxa    = clip_reclassify_align(JAXA_SEA, tile_boundary)

    # 3. Base agreement
    molca = where(fromglc == gl30, fromglc, 0)

    # 4. Thematic evidence
    molca = apply_builtup_evidence(molca, ghsbu, wsf)
    molca = apply_water_wetland_evidence(molca, gsw)
    molca = apply_forest_evidence(molca, fnf)

    # 5. Regional adaptation
    molca = apply_mangrove_rule(molca, namria, jaxa)
```

```
molca = fill_regional_agreement(molca, namria, jaxa)

# 6. Save outputs and reports
write_final_molca_tile(tile_id, molca)
write_nan_reduction_report(tile_id, molca)
```

The pseudocode is not a replacement for the final script. It is included to help the reader understand the logic. The complete implementation is documented separately in the project code and supporting appendices; the pseudocode here is limited to the scientific processing logic needed to understand the reference-map generation.

3.18 Boundary with the Later Shrubland Rescue Experiment

The workflow described in this chapter corresponds to the standard MOLCA Philippines reference-map generation stage. A later diagnostic phase showed that Shrubland required additional attention. That later phase produced a separate set of outputs stored under the Shrubland rescue folder, including new MOLCA variants and additional CSV reports. Those materials are not part of the standard MOLCA generation described in this chapter.

This separation is important for the chronology of the thesis. The standard MOLCA Philippines reference map was generated first. The deep learning model was then trained and evaluated. Only after the model diagnostics showed that Shrubland was almost not learned did we investigate whether the problem was caused only by the model or also by the reference-label construction. The Shrubland rescue therefore belongs mainly to the later experimental and results discussion, where the baseline model, WeightedRandomSampler test, confusion matrices, and CC-filtered reference variants are analysed together.

For this reason, Chapter 3 documents the standard reference-map generation workflow, while Chapter 6 discusses the evidence that motivated the later Shrubland rescue and evaluates its consequences.

3.19 Outputs of Chapter 3

The outputs of the Chapter 3 workflow are:

1. harmonised and aligned tile-level input rasters;
2. final MOLCA Philippines tile outputs;

3. visual quicklook maps for selected tiles;

4. staged AnA reduction CSV reports;

5. validation spreadsheets for logic checking;

6. a documented target legend for later patch selection and model training.

The final MOLCA outputs are not the final deep learning predictions. They are reference labels. Their function is to provide reliable training and validation information for the segmentation pipeline developed later in the thesis.

3.20 Limitations of the Reference-Map Generation

The MOLCA-derived reference map has several limitations. First, it depends on the quality of the input products. Agreement among products increases confidence, but it does not guarantee that a label is always correct. Two products can agree because they share similar input data, similar classification assumptions, or similar errors.

Second, the harmonisation process simplifies source legends. This is necessary, but it can remove thematic detail. For example, annual crop and perennial crop are both mapped to Cropland. Closed and open forest are both mapped to Forest. Fishpond is mapped to Water. These decisions are appropriate for the target legend, but they should be remembered when interpreting the final classes.

Third, AnA remains in the final map. This is intentional, but it means that the reference map is not a complete wall-to-wall national land-cover product. It is an agreement-derived reference layer designed for training and evaluation.

Fourth, the national and regional products improve local adaptation but introduce temporal differences. The workflow combines products from different nominal years. The thesis is not a land-cover change study, so the purpose is not to measure change between those years. The purpose is to construct a reliable reference map from the best available products. Still, temporal mismatch is a source of uncertainty and should be acknowledged.

Fifth, rare classes can remain difficult even after regional adaptation. Mangrove is strengthened by the national/regional agreement sources, but classes such as Shrubland may still be sensitive to source definitions and agreement thresholds. The model-based consequences of this issue are discussed later in Chapter 6.

3.21 Chapter Summary

This chapter described the generation of the MOLCA Philippines reference map. The method follows the agreement-based logic of MOLCA while adapting it to the Philippine study area. The workflow uses eight input products: FROM-GLC 2017, GlobeLand30 2020, GHS-BUILT-S, WSF 2019, GSW 2019, JAXA FNF, NAMRIA 2020, and JAXA Southeast Asia HRLULC 2023.

The procedure begins with tile-level preprocessing. Each source product is clipped, aligned, and reclassified to the common target legend. The base agreement is computed from FROM-GLC and GlobeLand30. Thematic evidence then reinforces Built-up, Water/Wetland, and Forest. NAMRIA and JAXA Southeast Asia are used to retain Mangrove as a separate class where both agree and to reduce remaining AnA where the two harmonised products agree on the same valid target class.

The staged AnA reduction reports show that the method substantially increases classified pixels while keeping uncertain locations unresolved. Four example tiles show final NaN percentages between 5.02% and 8.49%, compared with base-agreement NaN percentages between 60.43% and 72.44%. Validation spreadsheets further document that point-based rule checking was performed for selected tiles. The resulting reference map provides the label foundation for the patch selection, Sentinel-1 feature preparation, and deep learning segmentation chapters that follow.

Chapter 4

Patch Selection and Sentinel-1 Feature Preparation

4.1 Purpose of the Chapter

Chapter 3 described the generation of the MOLCA Philippines reference map. The next step was to convert this national-scale reference product into a practical training dataset for deep learning. This chapter explains how that transition was made. It describes the generation of candidate patches, the selection of the final 280 reference patches, the preparation of the Google Earth Engine table used for Sentinel-1 extraction, the 2019 Sentinel-1 processing strategy, and the local alignment between the SAR features and the MOLCA-derived ground-reference patches.

This chapter is important because it connects the reference-map construction of Chapter 3 with the deep learning methodology of Chapter 5. The model cannot be trained directly from the full national MOLCA raster. Instead, the data must be transformed into patch-level samples with matching input features and ground-reference labels. The goal is therefore to create a dataset in which each selected patch contains:

1. a MOLCA-derived ground-reference raster;
2. 28 Sentinel-1 VH feature rasters;
3. a consistent patch identifier;
4. matching spatial extent, size, and alignment between input and label;
5. metadata linking the Sentinel-1 files to the corresponding ground-reference patch.

The workflow described in this chapter was developed through an iterative process involving the project supervisors. Prof. Gamba indicated that the selected patches should be geographically sparse and should preserve, as much as possible, the class distribution of the full MOLCA-derived dataset. Tonia Sorriso then provided the operational guidance for the 20 × 20 patch grid, the final 549 × 549 patch size, the patch-selection criteria, and the Sentinel-1 temporal feature-extraction workflow. The selection stage produced 280 reference patches. After Sentinel-1 availability and completeness checks, three selected patches from tile 51QUC were excluded because they did not provide the complete required 2019 VH feature set. The final deep-learning dataset therefore contains 277 valid scene–reference associations.

This chapter is divided into two main parts. Sections 4.2–4.6 describe patch generation and selection. Sections 4.7–4.11 describe Sentinel-1 feature preparation and alignment.

4.2 Patch Candidate Generation

4.2.1 From full MOLCA tiles to candidate patches

The MOLCA Philippines reference map was generated tile by tile using the Sentinel-2 tiling framework. For deep learning, each tile had to be subdivided into smaller image patches. The selected patch size needed to be large enough to contain meaningful land-cover patterns, but small enough to be processed efficiently by the segmentation model and by the available hardware.

Following Tonia's guidance, each Sentinel-2 tile was partitioned into a regular 20 × 20 grid. This produces:

$$N_{\text{patches/tile}} = 20 \times 20 = 400. \tag{4.1}$$

The core Sentinel-2 tile size used in the patch-generation script was 10,980 pixels. Dividing this by 20 gives:

$$P_{\text{size}} = \frac{10980}{20} = 549. \tag{4.2}$$

Therefore, the final patch size was set to 549 × 549 pixels. This value is important because it became the common spatial unit for the MOLCA ground-reference patches and for the Sentinel-1 feature outputs. The final size also resolved the early ambiguity between the general idea of 512 × 512 patches and the actual CCI-style 20 × 20 grid used in the processing. In the final implementation, 549 × 549 is the size used for the prepared dataset.

For the 60 selected Sentinel-2 tiles, the full candidate pool is:

$$N_{\text{candidates}} = 60 \times 400 = 24000. \tag{4.3}$$

This candidate pool was not used directly. A filtering and selection procedure was needed because many patches were dominated by water, contained too much NoData/AnA, or did not contribute useful class diversity.

4.2.2 Patch-grid design

The patch grid was defined using row and column indices. Each patch identifier follows the form:

$$\text{patch\_id} = \text{TILE\_RXX\_CYY}, \tag{4.4}$$

where `TILE` is the Sentinel-2 tile name, `RXX` is the row index, and `CYY` is the column index. For example:

```
51PTR_R09_C15
```

identifies the patch located in row 09 and column 15 of tile 51PTR.

This naming convention is used throughout the workflow. It links the selected patch table, the extracted MOLCA ground-reference TIFF, the Sentinel-1 feature TIFFs, and the JSON metadata file. Keeping a consistent naming convention was essential because each selected patch ultimately corresponds to 28 separate Sentinel-1 feature rasters.

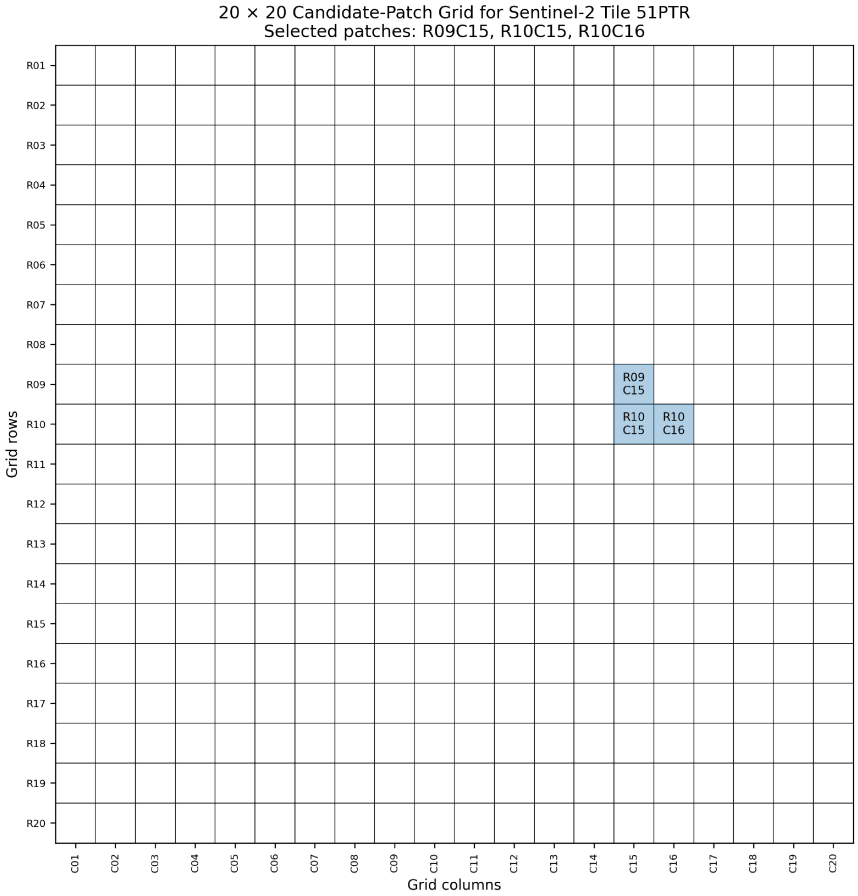


Figure 4.1: Schematic 20 × 20 candidate-patch grid for Sentinel-2 tile 51PTR. The highlighted cells correspond to the three patches selected from this tile: R09C15, R10C15, and R10C16. Each grid cell corresponds to an exact 549 × 549 pixel MOLCA reference patch.

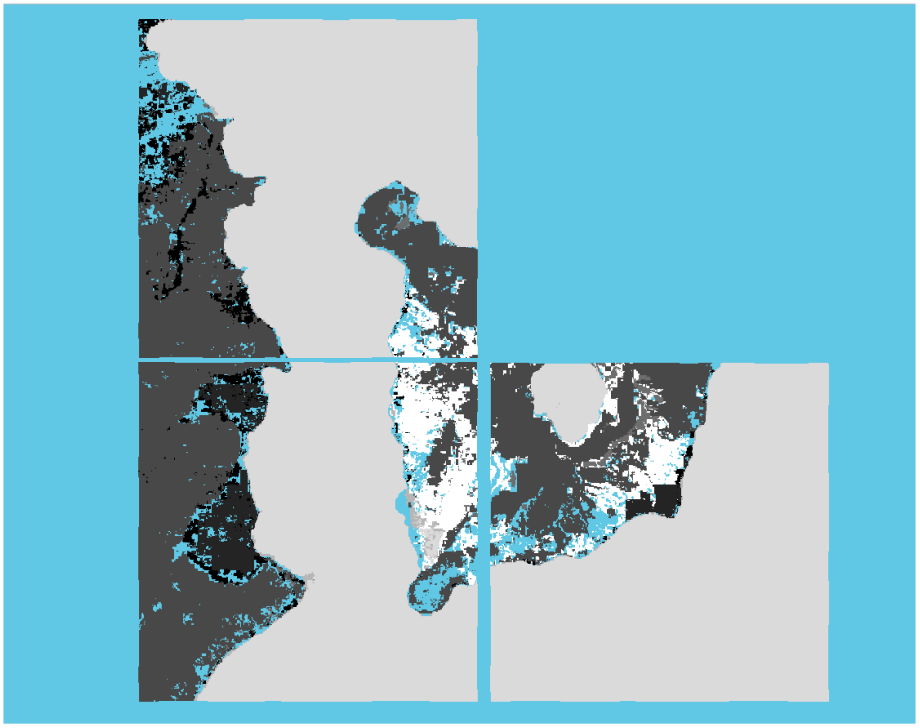


Figure 4.2: The three selected MOLCA-derived ground-reference patches from Sentinel-2 tile 51PTR. The arrangement corresponds to the grid positions highlighted in Figure 4.1. The figure shows the actual categorical reference information subsequently paired with the Sentinel-1 feature rasters.

4.3 Selection of the 60 Sentinel-2 Tiles

The first MOLCA processing stage was initially tested on a smaller subset of tiles. However, the analysis showed that a restricted 20-tile subset was not sufficient to represent the geographical and thematic diversity of the Philippines. The project was therefore expanded to 60 Sentinel-2 tiles. This decision improved the coverage of the country and increased the probability of including rare and spatially fragmented classes.

The final 60-tile set was used for both class-distribution analysis and candidate patch generation. The tile list was stored in `60_Tiles_name.csv`, while the final GEE table stored the selected patch extents. These 60 tiles define the operational area used for reference-patch selection and Sentinel-1 extraction. The later completeness check reduced the model-ready set from 280 selected patches to 277 valid scene–reference associations.

Table 4.1: Tile and candidate-patch structure used in Chapter 4.

Item	**Value / role**
Operational Sentinel-2 tiles	60 tiles.
Candidate grid per tile	20×20 grid.
Candidate patches per tile	400 patches.
Full candidate pool	24,000 candidate patches.
Final selected reference patches	280 patches.
Final patch size	549×549 pixels.
Model-ready associations after S1 completeness check	277 patches.
Patch identifier format	`TILE_RXX_CYY`.
Final patch table fields	`tile`, `patch_id`, `xmin`, `ymin`, `xmax`, `ymax`.
Coordinate system of patch extents	EPSG:4326 longitude/latitude.

4.4 Patch Selection Strategy

4.4.1 Selection objectives

The patch-selection strategy had to satisfy several constraints at the same time. A simple random selection would likely have produced many patches dominated by Water and Forest, because those classes occupy a large proportion of the valid MOLCA pixels. At the same time, forcing a perfectly balanced dataset would have produced an artificial distribution that no longer reflected the study area. The adopted strategy therefore aimed for a compromise between geographical coverage, class representativeness, and rare-class preservation.

The selection objectives were:

1. include all 60 operational tiles;
2. avoid selecting many neighbouring patches from the same area;
3. reject patches with too little valid labelled information;
4. preserve the overall class distribution of the MOLCA reference as much as possible;
5. reduce excessive dominance of pure Water patches;
6. ensure that rare classes such as Shrubland, Wetland, Bareland, and Mangrove were not completely missed;

7. produce a final dataset size suitable for Sentinel-1 feature extraction and model training.

This follows the guidance received from the supervisors: the selected subset should be geographically sparse and should resemble the distribution of labelled pixels in the full pseudo-MOLCA dataset, without severely underrepresenting less frequent classes.

4.4.2 Class-distribution analysis before selection

Before selecting the final patches, the class distribution of the 60-tile MOLCA reference was analysed. This step was necessary because the full reference map is highly imbalanced. Water and Forest are dominant, while Shrubland is extremely rare. The analysis showed that expanding from 20 to 60 tiles improved the representation of several classes, especially Grassland, Wetland, Bareland, and Mangrove, but did not solve the Shrubland scarcity.

Table 4.2 summarises the full 60-tile class counts recorded during the project workflow.

Table 4.2: Class distribution in the 60-tile MOLCA Philippines reference before final patch selection.

Class ID	**Class name**	**Pixels in 60-tile reference**
1	Built-up	71,247,316
2	Cropland	505,169,006
3	Forest	1,778,117,338
4	Grassland	117,238,811
5	Shrubland	16,699
6	Wetland	21,731,262
7	Water	3,279,717,034
8	Bareland	4,226,300
9	Mangrove Forest	20,388,898
–	Valid labelled pixels	5,797,852,664
–	NoData / AnA pixels	1,739,537,892
–	Total pixels	7,537,390,556

The table shows why patch selection required more than a random strategy. Water alone accounts for a very large portion of the valid pixels, followed by Forest and Cropland. Shrubland, by contrast, is almost absent in the standard MOLCA reference. This class imbalance later became relevant in the deep learning

experiments, but the first response to the problem was already present in Chapter 4: the patch-selection algorithm explicitly considered rare-class occurrence.

4.4.3 Rare-class tile analysis

To identify where rare and important classes were located, the 60-tile reference was analysed by class and by tile. The rare-class summary shows that Shrubland occurs in only six of the 60 tiles, while Wetland occurs in all 60, Bareland occurs in 59, and Mangrove occurs in 58. This confirms that not all rare classes have the same problem. Wetland and Mangrove are less common than Water or Forest, but they are geographically widespread. Shrubland is both extremely rare and spatially concentrated.

Table 4.3: Rare and important class presence across the 60 operational tiles.

ID	**Class**	**Tiles containing class**	**Pixels in 60 tiles**
4	Grassland	59	117,238,811
5	Shrubland	6	16,699
6	Wetland	60	21,731,262
8	Bareland	59	4,226,300
9	Mangrove Forest	58	20,388,898

The most important result is the Shrubland distribution. The top Shrubland tile is 51QTU, with 11,815 pixels, followed by 51PWP, 50PRC, 51QTV, 51QTA, and 51PTR. The remaining 54 tiles contain no Shrubland in the standard MOLCA output. This did not lead to an artificial over-selection of Shrubland patches, but it did justify giving Shrubland a higher priority score during patch selection.

Table 4.4: Top Shrubland tiles in the 60-tile MOLCA reference.

Rank	**Tile**	**Shrubland pixels**	**Share of valid tile pixels**
1	51QTU	11,815	0.010125%
2	51PWP	1,847	0.001645%
3	50PRC	1,763	0.001542%
4	51QTV	1,004	0.000855%
5	51QTA	173	0.000206%
6	51PTR	97	0.000085%

4.4.4 Selection scoring

The patch-selection scripts used a scoring strategy rather than simple random selection. The first stage generated exact 549×549 candidate patches and computed patch-level statistics. For each candidate patch, the script calculated the valid-pixel percentage, NoData/AnA percentage, class counts, class percentages, number of present classes, dominant class, and rare-class flags.

The selection then used a scoring function designed to reward useful patches and penalise low-information patches. The main criteria were:

1. valid-pixel percentage;
2. number of present land-cover classes;
3. presence of rare or important classes;
4. meaningful proportions of Wetland, Bareland, Mangrove, Shrubland, Grassland, and Built-up;
5. penalties for patches dominated by Water;
6. penalties for very low class diversity.

A simplified version of the refined selection score can be expressed as:

$$S_i = V_i + D_i + R_i + P_i - W_i - L_i, \tag{4.5}$$

where S_i is the score of candidate patch i, V_i is the valid-pixel contribution, D_i is the diversity score, R_i is the rare-class presence bonus, P_i is the capped class-proportion bonus, W_i is the water-dominance penalty, and L_i is the low-diversity penalty.

The exact implemented score was more detailed. It gave high weight to meaningful Shrubland, Wetland, Bareland, and Mangrove presence, while penalising patches where Water occupied most of the valid pixels. This design was necessary because Water is very dominant in the Philippine study area, but a training set made mostly of water patches would not support balanced land-cover learning.

4.4.5 Geographical constraint

The selection also enforced geographical spread. The refined selection script first selected a baseline number of patches per tile, then filled the remaining target size by global score while respecting a maximum number of patches per tile. This produced a dataset that includes every tile but still gives extra patches to tiles containing rare or useful classes.

The final CSV confirms that the selected dataset contains 280 patches across 60 tiles. Forty-eight tiles contribute three patches each, while selected rare-class or high-priority tiles contribute more. The highest-count tile is 51QTV with 24 patches, followed by 50PRC with 20, 51QTA with 16, and 51PTT and 51QTU with 14 each.

Table 4.5: Patch-count structure of the final 280-patch dataset.

Patch-count category	**Number of tiles**
3 selected patches	48 tiles
4 selected patches	1 tile
5 selected patches	1 tile
7 selected patches	2 tiles
8 selected patches	2 tiles
9 selected patches	1 tile
14 selected patches	2 tiles
16 selected patches	1 tile
20 selected patches	1 tile
24 selected patches	1 tile
Total selected patches	280
Represented tiles	60

This distribution reflects the hybrid strategy. The dataset remains geographically broad because all 60 tiles are represented, but it also allows selected tiles to contribute more patches when they contain rare or important land-cover conditions.

4.5 Class Distribution and Geographic Distribution

The final patch set was designed to represent both class and geography. These two objectives are not identical. A class-balanced sample could still be geographically clustered, while a geographically uniform sample could still miss rare classes. The final strategy therefore combines a minimum per-tile selection with score-based additional selection.

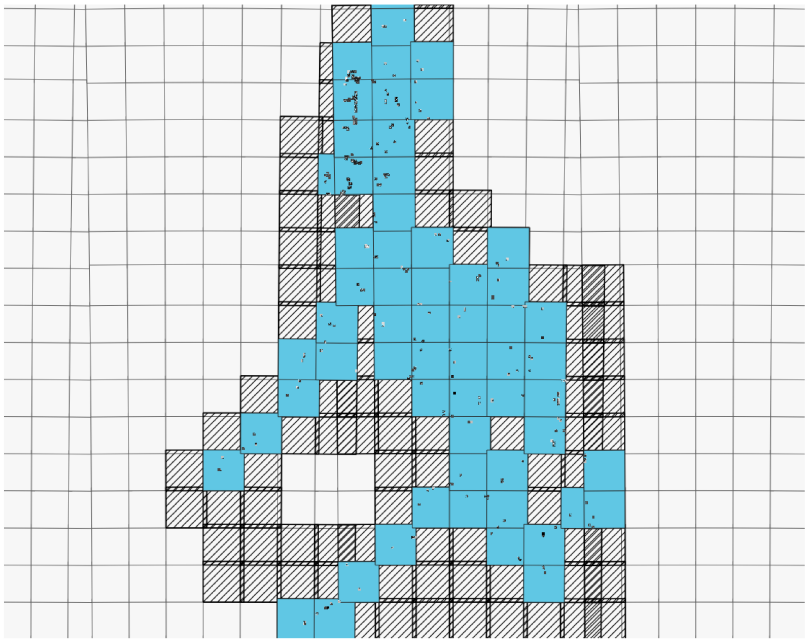
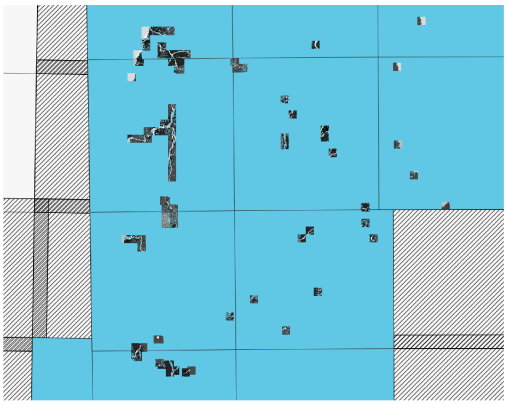
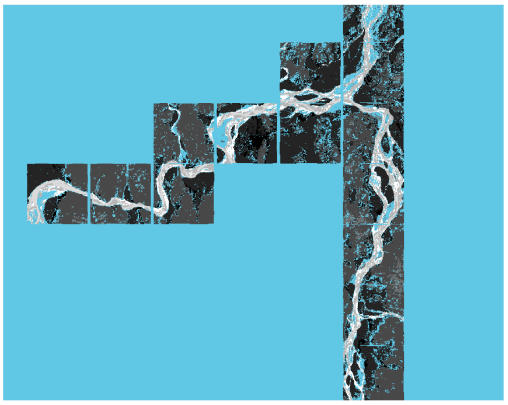


Figure 4.3: National-scale overview of the final 280 selected MOLCA-derived ground-reference patches across the 60 operational Sentinel-2 tiles. At this scale, the figure primarily illustrates the geographical coverage of the selected training areas across the Philippine study area.



(a) Intermediate-scale view showing a subset of the selected MOLCA ground-reference patches.



(b) Detailed view showing the spatial content of individual selected ground-reference patches.

Figure 4.4: Progressive spatial inspection of the final selected MOLCA-derived patch dataset. The two panels are enlarged views of the same 280-patch selection and illustrate how the national-scale sampling pattern translates into individual land-cover reference patches.

The geographical distribution supports the objective of training a model that is not limited to a small portion of the country. The selected patches include different island groups, coastal areas, inland landscapes, urban/peri-urban contexts, forests, agricultural zones, wetlands, and mangrove-rich areas. This is important because the later deep learning model must learn from a wide set of spatial contexts.

The class distribution objective was more difficult because the standard MOLCA reference is strongly imbalanced. In particular, the final selected patches could not make Shrubland abundant, because the class was almost absent in the standard 60-tile MOLCA reference. However, the selection algorithm was still useful because it captured a large fraction of the available Shrubland pixels without overrepresenting Water and Forest. This distinction is important: Chapter 4 improves the dataset selection within the available reference, but it cannot solve a reference-label scarcity problem by itself. The later Shrubland rescue work is discussed in Chapter 6.

4.6 Final 280 Selected Patches

4.6.1 Final GEE table

The final output of the patch-selection stage was a GEE-ready table containing all 280 selected reference patches:

`Final_GEE_Table_280_Patches.csv`

The table contains only the fields required for Sentinel-1 extraction:

`tile, patch_id, xmin, ymin, xmax, ymax`

Older intermediate fields such as patch width and patch height were intentionally ignored because they referred to intermediate extraction states. Tonia confirmed that this was correct, provided that the final patch table contained complete and unambiguous extents and that the final rasters would be brought to 549×549 during local post-processing.

Table 4.6: Final fields used for reconstructing patch geometries in Google Earth Engine.

Field	Meaning
`tile`	Sentinel-2 tile identifier.
`patch_id`	Unique patch identifier using the `TILE_RXX_CYY` convention.
`xmin`	Western longitude of the patch bounding box.
`ymin`	Southern latitude of the patch bounding box.
`xmax`	Eastern longitude of the patch bounding box.
`ymax`	Northern latitude of the patch bounding box.

The patch extents are stored in EPSG:4326 longitude/latitude coordinates. In GEE, each patch geometry is reconstructed directly from these bounding-box coordinates. This removed the need to upload a separate patch shapefile asset, although generating a geometry layer from the CSV would also have been possible.

4.6.2 Patch IDs and export naming

The final patch IDs are unique and suitable for export naming. This is important because each patch generates 28 Sentinel-1 feature files. A unique patch ID prevents confusion between different tiles and positions. It also allows a simple matching procedure between the Sentinel-1 folder, the ground-reference folder, and the label metadata.

For example, the patch:

```
51PTR_R09_C15
```

can be linked to ground-reference and Sentinel-1 features using the same tile-row-column identifier.

4.6.3 From 280 selected patches to 277 model-ready associations

The number of selected reference patches and the number of samples ultimately used for deep learning are not identical. The patch-selection stage produced 280 candidates for Sentinel-1 extraction. During the subsequent availability and completeness check, three selected patches from Sentinel-2 tile 51QUC did not provide the complete required set of 2019 descending-orbit VH features. Those three patches were retained in the selection history but excluded from the model-ready dataset.

Consequently, the project uses two related counts for different stages of the workflow:

- 280 selected MOLCA-derived reference patches for the planned Sentinel-1 extraction;
- 277 valid scene–reference associations with a complete 28-feature Sentinel-1 input set for deep learning.

This distinction is maintained throughout the later methodology and results chapters.

4.7 Sentinel-1 Data Collection for 2019

4.7.1 Sentinel-1 configuration

The predictor data used in this thesis are derived from Sentinel-1 SAR imagery for the year 2019. The GEE script uses the `COPERNICUS/S1_GRD` image collection and filters it by:

1. acquisition year: 2019;
2. orbit pass: descending;
3. instrument mode: IW;
4. polarisation: VH;
5. spatial intersection with the patch or tile processing geometry.

The Sentinel-1 configuration is summarised in Table 4.7.

Table 4.7: Sentinel-1 configuration used for the Philippine feature-extraction workflow.

Parameter	**Value used in the workflow**
GEE collection	`COPERNICUS/S1_GRD`
Year	2019
Orbit pass	Descending
Instrument mode	IW
Polarization	VH only
Patch-coordinate CRS	EPSG:4326
Export CRS from GEE	EPSG:4326
Export scale	10 m
Texture radius	30 m
Output features per patch	28 single-band GeoTIFFs

VH was used as the only polarisation in the final workflow. This choice follows the instructions received for the Sentinel-1 feature extraction and keeps the input representation consistent across all selected patches.

4.7.2 Patch-based extraction instead of full-tile export

The Sentinel-1 feature extraction was designed to be computationally efficient. Tonia recommended computing the seasonal feature stack at tile level and then cropping only the selected patches using the coordinates in the final table. This avoids recomputing the expensive seasonal feature generation for every patch independently.

The logic is:

1. read the selected patches for one tile;
2. reconstruct patch geometries from the CSV extents;
3. define the processing geometry for the tile or selected patch group;
4. filter the 2019 Sentinel-1 collection;
5. compute the seasonal feature stack once;
6. export each feature for each selected patch.

This design is efficient because the seasonal Sentinel-1 stack is shared among all selected patches in the same tile. Only the export/cropping step is repeated for each patch.

4.8 Seasonal Feature Extraction

4.8.1 Seasonal subdivision

The 2019 Sentinel-1 time series was divided into four seasonal subsets:

Table 4.8: Four temporal groups used to divide the Sentinel-1 2019 time series.

Temporal group	**Internal code**	**Date range**
Season 1	WI	1 January–31 March
Season 2	SP	1 April–30 June
Season 3	SU	1 July–30 September
Season 4	AU	1 October–31 December

The terms `WI`, `SP`, `SU`, and `AU` are retained as internal identifiers because they are used in the processing script and output filenames. In the thesis text, however, the four periods are referred to as Season 1–Season 4. This avoids assigning temperate-season terminology to the tropical Philippine study area while preserving correspondence with the implemented processing workflow.

In the implementation, GEE uses exclusive end dates in `filterDate()`. Therefore, the script uses 1 April, 1 July, 1 October, and 1 January 2020 as the exclusive end dates of temporal groups 1–4, respectively.

4.8.2 Super Image and denoising

For each season, the script builds a seasonal Super Image. The Super Image is the seasonal mean of the denoised Sentinel-1 images within that temporal subset. Before the seasonal mean is computed, the script applies a ratio-based multitemporal despeckling step. It first constructs an annual Super Image from the full 2019 collection. Each image is converted from decibel to natural units, divided by the annual Super Image, filtered using the Refined Lee method, multiplied back by the annual Super Image, and converted back to decibels.

In simplified notation, let $I_k(p)$ be the Sentinel-1 VH backscatter image at time k for pixel p, expressed in natural units, and let $\bar{I}_Y(p)$ be the annual mean image. The denoised image can be represented conceptually as:

$$\tilde{I}_k(p) = \mathcal{L}\left(\frac{I_k(p)}{\bar{I}_Y(p)}\right)\bar{I}_Y(p), \tag{4.6}$$

where $\mathcal{L}(\cdot)$ represents the Refined Lee filtering operation. This adaptive filter is used to reduce speckle while preserving local structural information more effectively than simple uniform smoothing [34]. In the implemented Earth Engine

workflow, a 7×7 directional neighbourhood is used to estimate local statistics and determine the filtered backscatter value.

The seasonal Super Image for season s is then defined as:

$$SI_s(p) = \frac{1}{N_s} \sum_{k \in s} \tilde{I}_k(p), \tag{4.7}$$

where N_s is the number of Sentinel-1 images available in season s.

The purpose of this step is to reduce speckle while preserving the temporal and spatial structure of the SAR signal. In the thesis workflow, the Super Image is not the final output by itself; it is also the base layer from which local spatial features are derived.

4.8.3 Seven seasonal features

For each season, seven features are produced:

1. SI: seasonal Super Image;
2. LEE: Lee-filtered seasonal Super Image;
3. MAX: local maximum;
4. MIN: local minimum;
5. MAXMIN: local maximum minus local minimum;
6. MEAN: local mean;
7. MEDIAN: local median.

The local statistics are computed using a square kernel with a 30 m radius. If $K(p)$ is the neighbourhood around pixel p, then the local features are:

$$MAX_s(p) = \max_{q \in K(p)} SI_s(q), \tag{4.8}$$

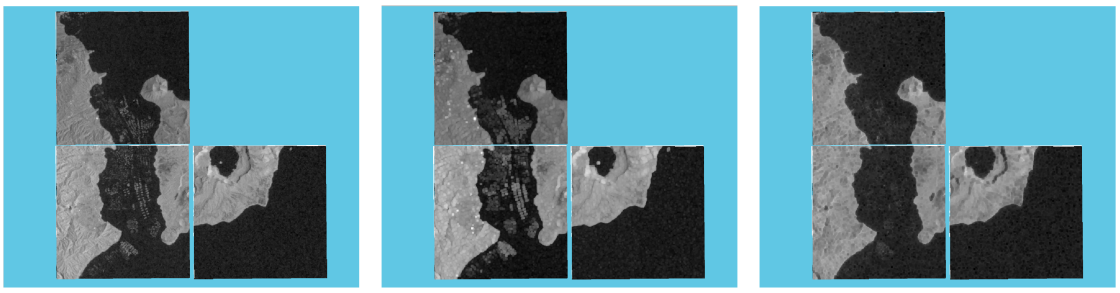
$$MIN_s(p) = \min_{q \in K(p)} SI_s(q), \tag{4.9}$$

$$MAXMIN_s(p) = MAX_s(p) - MIN_s(p), \tag{4.10}$$

$$MEAN_s(p) = \frac{1}{|K(p)|} \sum_{q \in K(p)} SI_s(q), \tag{4.11}$$

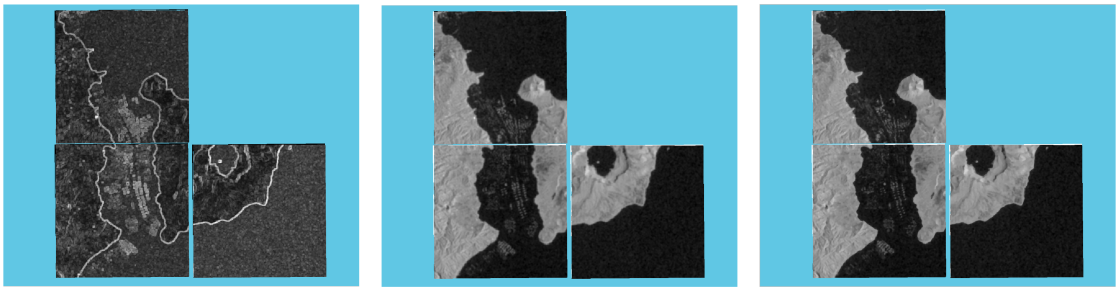
$$MEDIAN_s(p) = \text{median}_{q \in K(p)} SI_s(q). \tag{4.12}$$

Together, these features describe both seasonal radar backscatter and local spatial texture. The resulting representation is more informative than a single Sentinel-1 image because it captures temporal aggregation and local neighbourhood structure.



(a) LEE (b) MAX (c) MIN

Figure 4.5: Examples of three Sentinel-1 features derived from the first seasonal Super Image for the selected 51PTR patches: Lee-filtered SI, local maximum, and local minimum.



(a) MAXMIN (b) MEAN (c) MEDIAN

Figure 4.6: Additional Sentinel-1 spatial features derived from the first seasonal Super Image: local range (MAXMIN), local mean, and local median. Together with SI, LEE, MAX, and MIN, these form the seven features generated for each season.

4.9 Generation of Sentinel-1 Feature Stacks

4.9.1 The 28-feature representation

The Sentinel-1 representation contains 28 features per selected patch. This is obtained by applying the seven-feature extraction procedure to each of the four temporal groups:

$$N_{\text{features}} = N_{\text{groups}} \times N_{\text{features/group}} = 4 \times 7 = 28. \tag{4.13}$$

Table 4.9 shows the feature plan.

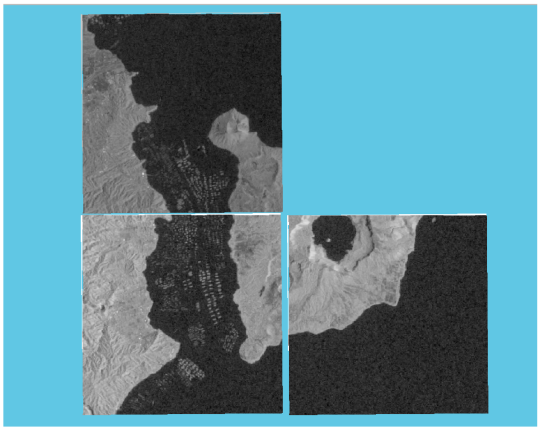
Table 4.9: Sentinel-1 feature plan used for every selected patch.

Temporal group	**Code**	**Features**
Season 1	01 / WI	SI, LEE, MAX, MIN, MAXMIN, MEAN, MEDIAN
Season 2	02 / SP	SI, LEE, MAX, MIN, MAXMIN, MEAN, MEDIAN
Season 3	03 / SU	SI, LEE, MAX, MIN, MAXMIN, MEAN, MEDIAN
Season 4	04 / AU	SI, LEE, MAX, MIN, MAXMIN, MEAN, MEDIAN
Total	–	28 features per patch

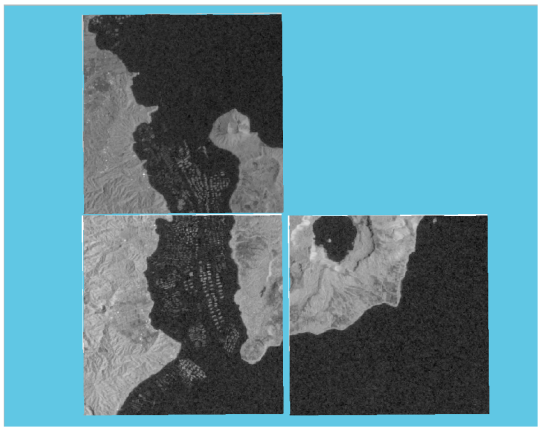
For the 280 selected patches, the planned extraction workload was:

$$N_{\text{S1 files}} = 280 \times 28 = 7840. \tag{4.14}$$

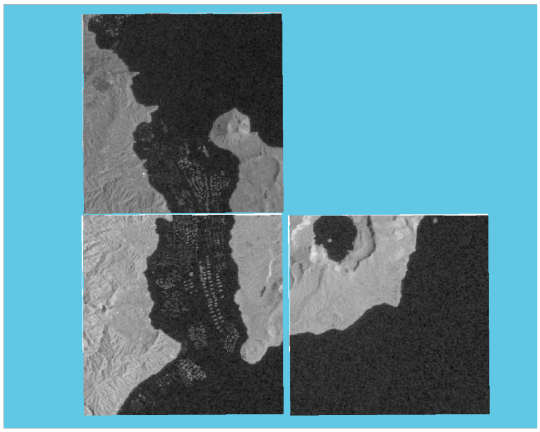
This value matches the processing checklist and the completed raw extraction campaign: 7,840 feature-export tasks were generated for the original 280-patch selection. The later model-readiness check is distinct from this export count; three selected 51QUC patches were subsequently excluded because a complete valid 2019 VH feature set was not available, leaving 277 usable scene–reference associations for deep learning.



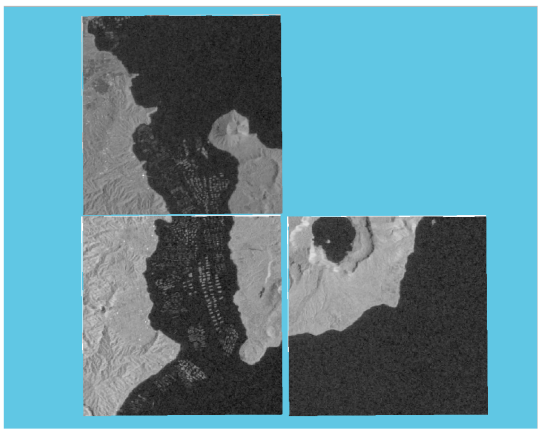
(a) Season 1: January–March 2019.



(b) Season 2: April–June 2019.



(c) Season 3: July–September 2019.



(d) Season 4: October–December 2019.

Figure 4.7: Seasonal Sentinel-1 VH Super Images for the three selected patches of tile 51PTR. The four panels correspond to the four temporal groups used in the feature-extraction workflow. Although the spatial structures remain geographically consistent, the SAR response varies across the year, providing temporal information that complements the spatial features extracted from each seasonal Super Image.

Figure 4.7 provides a visual example of the temporal component of the Sentinel-1 input representation. The same three geographical patches are shown for all four temporal groups, allowing differences in the seasonal VH response to be inspected while preserving the same spatial context. The grayscale rendering is used only for visual interpretation in QGIS; the deep learning pipeline uses the original floating-point raster values rather than the displayed grayscale values.

For every temporal group, the seasonal Super Image is subsequently used to derive the seven representations described in Section 4.8.3. Therefore, temporal

variability and local spatial information are combined within the final 28-feature input representation.

4.9.2 Export naming convention

The GEE script exports each Sentinel-1 feature as a separate single-band GeoTIFF. The naming convention combines the patch ID, year, season code, feature code, and feature label:

```
patchID_YEARseasonCodefeatureCode_FEATURE.tif
```

For example:

```
51PTR_R09_C15_20190101_SI.tif
51PTR_R09_C15_20190102_LEE.tif
51PTR_R09_C15_20190103_MAX.tif
...
```

During local post-processing, the naming convention is converted to the structure requested for the training dataset:

```
TILE_YEARSEASONFEATUREID_FEATURE_ROW_COL.tif
```

For example:

```
51PTR_20190101_SI_09_15.tif
51PTR_20190102_LEE_09_15.tif
```

This consistent naming structure is essential because the deep learning loader must associate each ground-reference patch with its corresponding 28 Sentinel-1 input rasters.

4.10 Google Earth Engine Implementation

4.10.1 CSV asset preparation

The final patch table was uploaded to Google Earth Engine as a table asset. The GEE script reads this asset and reconstructs patch geometries using the bounding-box columns. The CSV had to be saved as a clean comma-delimited file because an early version contained a Byte Order Mark that caused Earth Engine to misread the first column name. After re-saving the CSV and making the geometry-building function more robust, the table could be read correctly.

The key asset configuration was:

```
PATCH_CSV_ASSET
TILE_COLUMN = 'tile'
PATCH_ID_COLUMN = 'patch_id'
XMIN_COLUMN = 'xmin'
YMIN_COLUMN = 'ymin'
XMAX_COLUMN = 'xmax'
YMAX_COLUMN = 'ymax'
PATCH_COORDS_CRS = 'EPSG:4326'
```

This approach avoided the need to upload the patch polygons separately. The geometry of each patch was reconstructed directly from the CSV.

4.10.2 Single-tile and single-patch testing

Before launching the full extraction, the script was tested on tile 51PTR and on a small number of patches. This test was necessary to check four elements:

1. whether the patch geometries were reconstructed correctly;
2. whether the Sentinel-1 collection contained valid 2019 VH images;
3. whether the exported TIFFs had real radar values;
4. whether QGIS could display the exported rasters after proper stretching.

The first exported images appeared black in QGIS. This was not necessarily a data error. Sentinel-1 backscatter values are not 8-bit RGB image values; they often need proper display stretching. After adjusting QGIS symbology using actual min/max or standard-deviation stretching, the grayscale radar structure became visible.

A second issue was the coverage filter. The original script contained a minimum tile-coverage threshold. For patch-based extraction, this filter removed valid Sentinel-1 images because the selected patches are much smaller than a full Sentinel-2 tile. After diagnostic checking, the coverage filter was disabled through a flag, following Tonia's recommendation. This retained the option of using the filter later for full-tile processing while allowing patch-based extraction to proceed correctly.

Table 4.10: Main technical issues encountered during the GEE Sentinel-1 test and how they were handled.

Issue	**Observed problem**	**Resolution**
CSV encoding	GEE did not correctly read the first column name due to a BOM issue.	The CSV was re-saved as a clean comma-delimited file, and robust string handling was added in the geometry-building function.
Black QGIS display	Initial Sentinel-1 exports appeared black in QGIS.	QGIS symbology was adjusted using real raster min/max or standard-deviation stretching.
Coverage threshold	The full-tile coverage filter excluded valid patch-level scenes.	A flag was introduced to enable or disable the coverage filter; for patch extraction it was disabled.
Patch dimensions in EPSG:4326	GEE exports in EPSG:4326 did not always display as exactly 549 × 549 pixels.	Final reprojection, snapping, and cropping were handled locally after download.

4.10.3 Batch automation

The raw extraction campaign involved 7,840 Sentinel-1 feature exports for the 280 selected patches. This required automation because manually launching and monitoring every task would be inefficient. After the initial JavaScript/GEE tests, the same processing logic was executed through the Earth Engine Python API in manageable tile batches. The runner tracked active Earth Engine tasks and submitted additional batches as capacity became available. The processing checklist records completion of the 60-tile extraction campaign and the expected 7,840 raw feature exports.

The automation script also included optional Telegram notifications for long-running batches. Authentication tokens and chat identifiers are operational credentials and are not part of the scientific workflow; they must therefore be omitted or redacted from any appendix or shared code archive.

4.11 Alignment between Reference Patches and Sentinel-1 Features

4.11.1 Need for local alignment

The GEE exports were produced in EPSG:4326 at 10 m scale. This was acceptable for extraction, but the final deep learning dataset required the Sentinel-1 feature rasters and MOLCA ground-reference rasters to match exactly in pixel size and spatial extent. Therefore, a local GDAL-based post-processing step was used to reproject, snap, crop, and rename the files.

This step is critical. Even small differences between input rasters and label rasters can create wrong pixel-to-label correspondence during segmentation training. The model assumes that each input pixel and its target label refer to the same ground location. If the SAR feature and the ground-reference patch are shifted or have different dimensions, the training labels become spatially inconsistent.

4.11.2 Ground-reference preparation

The local post-processing script first processes the MOLCA ground-reference patch. Each ground-reference TIFF is reprojected and forced to 549×549 pixels using nearest-neighbour resampling. Nearest-neighbour resampling is required for categorical labels because the MOLCA class IDs must not be interpolated.

The ground-reference naming convention is:

```
TILE_GT_ROW_COL.tif
```

For example:

```
51PTR_GT_09_15.tif
```

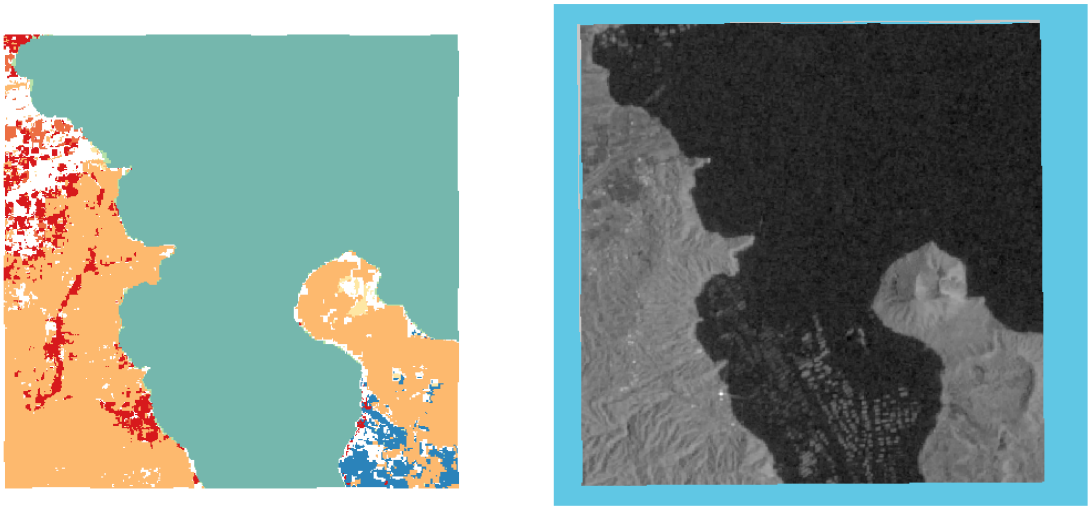
The ground-reference patch is then used as the spatial template for the corresponding Sentinel-1 features.

4.11.3 Sentinel-1 feature alignment

After the ground-reference patch is prepared, the script reads its projected bounding box and uses it to align the corresponding Sentinel-1 feature rasters. Unlike the categorical ground-reference raster, the Sentinel-1 rasters contain continuous backscatter-derived values. Therefore, bilinear resampling is used for the SAR features.

The local alignment procedure can be summarised as:

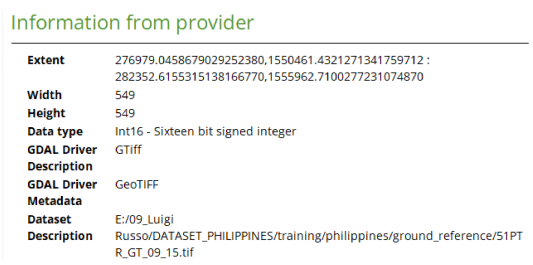
1. reproject the ground-reference patch;
2. force the ground-reference raster to 549 × 549 pixels;
3. extract the projected bounding box of the ground-reference patch;
4. find all 28 matching Sentinel-1 feature files;
5. reproject and crop each Sentinel-1 feature to the ground-reference bounding box;
6. force each Sentinel-1 feature to 549 × 549 pixels;
7. save the aligned files using the final naming convention;
8. create a JSON metadata file linking the ground-reference patch to its 28 Sentinel-1 rasters.



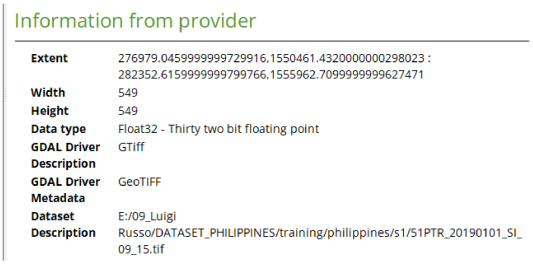
(a) MOLCA-derived ground-reference patch.

(b) Corresponding Sentinel-1 Season 1 SI raster.

Figure 4.8: Visual verification of the correspondence between the MOLCA-derived ground-reference patch and its Sentinel-1 input raster for patch 51PTR_R09_C15 after local post-processing. The ground-reference layer is rendered using categorical pseudocolour, while the Sentinel-1 raster is displayed in grayscale because it contains continuous backscatter-derived values. Both panels represent the same geographical footprint.



(a) Ground-reference raster information.



(b) Sentinel-1 SI raster information.

Figure 4.9: QGIS raster-information panels used to verify the result of local post-processing for patch 51PTR_R09_C15. Both the categorical ground-reference raster and the continuous Sentinel-1 SI raster have dimensions of 549 × 549 pixels and virtually identical projected extents. The very small numerical differences in the displayed extent coordinates are attributable to floating-point representation and are negligible relative to the 10 m spatial resolution.

The alignment verification shown in Figure 4.9 provides a direct numerical check of the local post-processing procedure. Both rasters have a width and height of 549 pixels. The ground-reference raster is stored as an integer categorical layer, whereas the Sentinel-1 feature is stored as a floating-point raster, as expected from their different roles in the segmentation dataset.

The reported extents are also effectively coincident at the mapping scale of this dataset. The small numerical differences visible in the QGIS metadata panels are negligible relative to the 10 m pixel spacing. For the verification example shown here, the input SAR feature and its corresponding categorical reference label therefore occupy the same spatial footprint after local alignment.

The alignment was verified at two complementary levels. First, the ground-reference and Sentinel-1 layers were visually compared in QGIS to confirm correspondence of coastlines, terrain structures, and other spatial features. Second, their raster metadata were inspected to verify identical 549 × 549 dimensions and matching projected extents. This combination of visual and numerical verification provides stronger evidence than relying on either inspection alone.

4.11.4 JSON metadata

For each patch, a JSON file is created to store the relationship between the ground-reference patch and the corresponding Sentinel-1 feature files. The JSON structure includes:

```
corresponding_s1
labels
```

The `corresponding_s1` field lists the 28 aligned Sentinel-1 feature files associated with the patch. The `labels` field records the land-cover class IDs present in the ground-reference patch. This metadata supports later dataset loading and quality checking.

The final folder structure is:

```
Final_Output/
    Ground_Reference/
    Labels/
    S1/
```

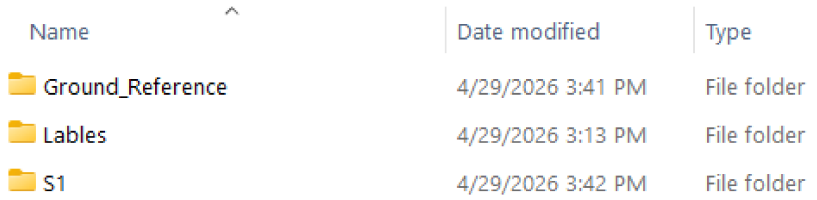


Figure 4.10: Final dataset organisation after local post-processing. The folder structure separates Sentinel-1 input features, MOLCA ground-reference patches, and JSON label metadata.

4.12 Chapter 4 Workflow Summary

Figure 4.11 summarises the complete Chapter 4 workflow.

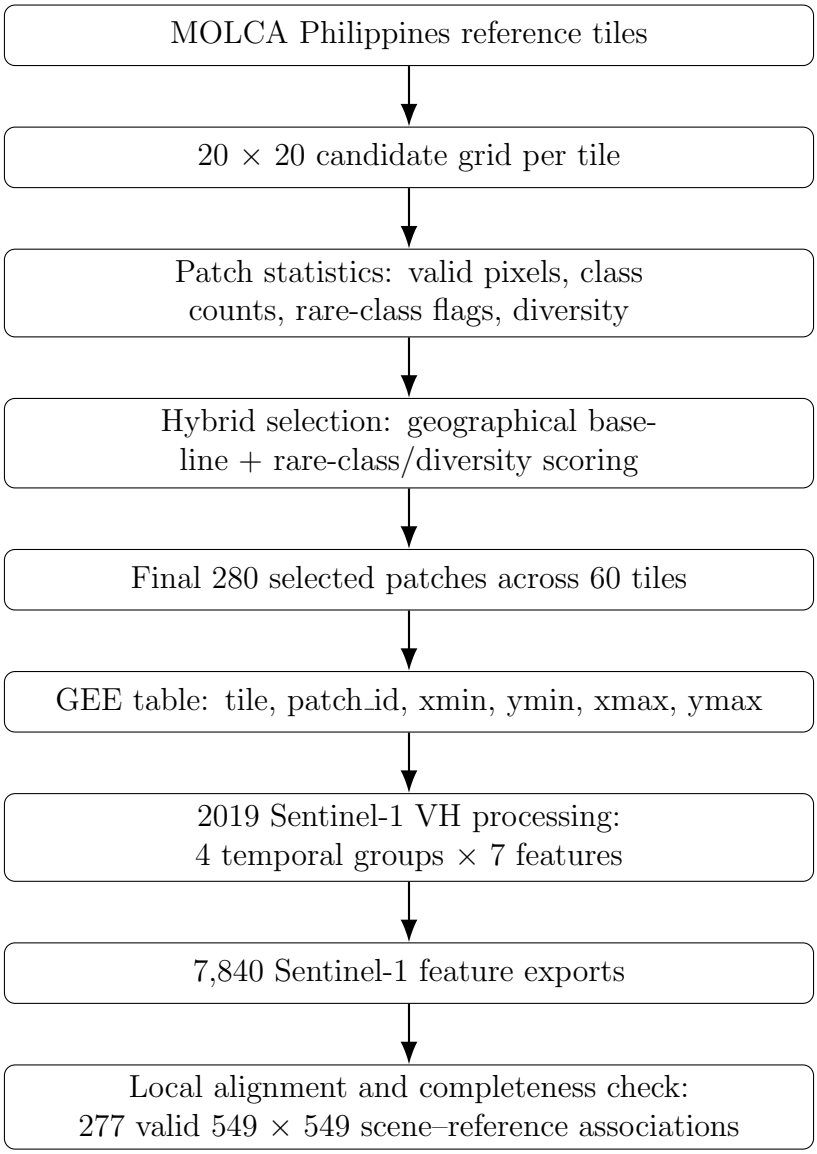


Figure 4.11: Complete workflow from MOLCA reference tiles to the 280-patch Sentinel-1 extraction and the final 277 aligned, model-ready scene–reference associations.

4.13 Limitations and Practical Notes

Several limitations should be noted.

First, the selected 280 patches are not a random sample of the Philippines. They are a designed training subset. This is appropriate for deep learning prepa-

ration, but the selection should not be interpreted as an independent statistical sample for estimating national land-cover proportions.

Second, the patch-selection strategy improves rare-class representation only within the limits of the standard MOLCA reference. It cannot create reliable Shrubland labels where the reference map contains almost none. This explains why Shrubland remains a later diagnostic issue in the thesis.

Third, the GEE exports in EPSG:4326 were not the final training rasters. They were intermediate outputs. The final training-ready files required local alignment, reprojection, cropping, and snapping to match the ground-reference patches.

Fourth, the final model-ready dataset contains 277 valid scene–reference associations rather than all 280 selected reference patches. The three excluded 51QUC patches did not provide the complete required 2019 VH feature set and were therefore not passed to model training.

Fifth, the extraction automation contained private notification credentials. These are unrelated to the scientific method and must be redacted from any appendix or public code release.

4.14 Chapter Summary

This chapter described how the MOLCA Philippines reference map was converted into a patch-level dataset suitable for deep learning. The workflow began with the 60 selected Sentinel-2 tiles. Each tile was divided into a 20 × 20 candidate grid, producing 24,000 candidate patches. The final patch size was set to 549 × 549 pixels.

The patch-selection process combined geographical coverage with class-aware scoring. All 60 tiles were represented, while additional patches were assigned to tiles containing rare or important classes. The selection stage produced 280 reference patches, stored in a GEE-ready CSV table with the fields `tile`, `patch_id`, `xmin`, `ymin`, `xmax`, and `ymax`. After Sentinel-1 completeness checks, three 51QUC patches were excluded, leaving 277 valid scene–reference associations for the deep-learning stage.

The Sentinel-1 feature preparation used 2019 descending-orbit VH imagery from the `COPERNICUS/S1_GRD` collection. The time series was divided into four fixed temporal groups corresponding to January–March, April–June, July–September, and October–December. For each group, seven features were produced: SI, LEE, MAX, MIN, MAXMIN, MEAN, and MEDIAN. This defines 28 Sentinel-1 features per patch. The original 280-patch extraction therefore comprised 7,840 raw feature exports; the final model-ready dataset contains 277 complete scene–reference associations.

Finally, the exported Sentinel-1 rasters and MOLCA-derived ground-reference patches were locally post-processed to produce spatially corresponding training samples. The ground-reference rasters were treated as categorical data, while the Sentinel-1 features retained their continuous floating-point representation. A verification example for 51PTR_R09_C15 confirmed identical 549 × 549 raster dimensions and effectively coincident projected extents between the ground-reference patch and its corresponding Sentinel-1 SI raster. JSON metadata were then used to link each ground-reference patch to its associated Sentinel-1 feature files and land-cover labels. The resulting dataset forms the direct input to the deep learning methodology described in Chapter 5.

Chapter 5

Deep Learning Classification Methodology

The previous chapters described the construction of the MOLCA-derived reference map for the Philippines, the selection of representative training areas, and the preparation of the corresponding Sentinel-1 (S1) synthetic aperture radar (SAR) feature stacks. Once the reference and SAR inputs had been brought into a common spatial framework, the remaining task was to formulate the land-cover mapping problem as a supervised dense classification problem and to train a model capable of assigning a land-cover label to every valid pixel.

This chapter describes that deep learning stage in detail. Particular attention is given to the exact implementation used in this thesis because several elements of the original processing chain were adapted to the Philippine case. The general methodological basis follows the S1 land-cover mapping framework proposed by Russo et al. [9], which combines seasonally synthesised SAR features with a transformer-based U-shaped segmentation network. However, the implementation used here is not a direct reproduction of every experimental choice reported in that work. The study area, reference map, acquisition year, land-cover legend, training set, class distribution, model implementation, training duration, imbalance handling, and evaluation procedure were adapted to the dataset developed in Chapters 3 and 4.

The final deep learning dataset contains 277 valid scene samples. Each scene is associated with one MOLCA-derived reference raster and the complete sequence of 28 Sentinel-1 feature rasters used by the final model. The 28 features correspond to seven spatial or statistical representations computed for each of four temporal groups. Before being presented to the model, each scene is resized to 512×512 pixels and subdivided into nine overlapping 256×256 samples with a stride of 128 pixels. The network therefore receives a tensor with 28 input channels and produces a dense ten-channel output corresponding to the ten-class Philippine

legend.

An important aspect of the methodology is that the train–validation separation is carried out at the *scene level*, before the overlapping 256×256 samples are generated. Consequently, overlapping samples originating from one 512×512 scene cannot be divided between training and validation. This design is particularly important for remote-sensing segmentation because randomly splitting strongly overlapping image chips can lead to spatial leakage and an overly optimistic estimate of model generalisation.

The experiments were also designed to address the strong class imbalance found in the reference data. Two complementary mechanisms were considered. First, a class-weighted cross-entropy loss increases the contribution of under-represented classes at the pixel level. Second, a controlled experiment based on a weighted random sampler increases the frequency with which training patches containing Shrubland are presented to the model. These two mechanisms operate at different levels and are therefore described separately in this chapter.

The numerical outcomes of the training experiments are intentionally not reported here. Chapter 6 presents the training curves, confusion matrices, class-wise scores, prediction maps, the comparison between the clean CC50 baseline and the Shrubland-focused sampling experiment, and the subsequent interpretation of the remaining errors.

5.1 Semantic Segmentation Framework

5.1.1 From land-cover classification to dense semantic segmentation

Land-cover mapping can be approached at different spatial levels. A conventional image-classification model assigns one label to an entire image or image patch. Such a formulation is unsuitable for the present problem because each selected Philippine scene generally contains several land-cover types. A single 512×512 scene may simultaneously contain Forest, Cropland, Water, Built-up areas, Grassland, Wetland, or other classes. Assigning one label to the whole scene would therefore discard the spatial structure that the reference map was specifically designed to preserve.

The problem is instead formulated as *semantic segmentation*. Semantic segmentation assigns a semantic category to every pixel of an image while exploiting the surrounding spatial context to make that assignment. U-shaped convolutional architectures such as U-Net established an effective encoder–decoder design for dense segmentation by combining context extracted at progressively coarser scales with fine spatial information transmitted through skip connections [13].

Transformer-based variants extend this principle by allowing relationships over larger spatial contexts to be represented through attention mechanisms.

Let an input sample be represented by

$$\mathbf{X} \in \mathbb{R}^{C \times H \times W}, \tag{5.1}$$

where C denotes the number of input channels and H and W denote the spatial dimensions. In this thesis,

$$C = 28, \qquad H = W = 256 \tag{5.2}$$

during network training. The segmentation network, parameterized by θ, implements a mapping

$$f_\theta : \mathbb{R}^{28 \times 256 \times 256} \to \mathbb{R}^{10 \times 256 \times 256}. \tag{5.3}$$

For a mini-batch of size B, the corresponding dimensions are

$$\mathbf{X} \in \mathbb{R}^{B \times 28 \times 256 \times 256} \tag{5.4}$$

and

$$\mathbf{Z} = f_\theta(\mathbf{X}) \in \mathbb{R}^{B \times 10 \times 256 \times 256}, \tag{5.5}$$

where $\mathbf{Z}$ contains the raw class scores, or logits.

For pixel i and class c, the softmax probability is

$$p_{i,c} = \frac{\exp(z_{i,c})}{\sum_{j=0}^{K-1} \exp(z_{i,j})}, \qquad K = 10. \tag{5.6}$$

The predicted class is obtained from the maximum posterior probability,

$$\hat{y}_i = \arg\max_c p_{i,c}. \tag{5.7}$$

Although the prediction is finally expressed as one class per pixel, the decision is not made from that pixel in isolation. The encoder–decoder network uses contextual information surrounding each pixel and represents increasingly large spatial structures as information passes through the hierarchy of the model. This is particularly useful in SAR-based land-cover classification, where different land-cover types may exhibit overlapping backscatter characteristics at the individual-pixel level.

5.1.2 Why dense segmentation is appropriate for SAR land-cover mapping

The spatial organisation of land cover contains information that cannot be represented by a purely per-pixel classifier. Forest, agricultural areas, rivers, coastlines, settlements, wetlands, and mangrove environments differ not only in their SAR backscatter values but also in their spatial extent, texture, boundaries, fragmentation, and relation to adjacent classes. The 28-channel representation developed in Chapter 4 provides the model with seasonal and spatial SAR descriptors, while semantic segmentation allows those descriptors to be interpreted together with their spatial context.

This point is particularly relevant for the Philippine study area. The archipelagic geography produces frequent coastlines and mixed land–water boundaries, while tropical vegetation produces several classes with potentially similar radar responses. A dense segmentation architecture can combine fine-scale evidence with wider contextual information. For example, a local backscatter pattern may be ambiguous when considered alone, but its relation to a larger forest block, river corridor, coastline, or agricultural mosaic may provide additional information for classification.

The objective of the model is therefore not merely to associate a backscatter value with a class. Instead, the model learns a nonlinear mapping from a 28-dimensional spatio-temporal SAR representation and its surrounding context to the corresponding semantic land-cover category.

5.1.3 Relationship with the methodology of Russo et al.

The deep learning component of this thesis originates from the land-cover classification framework developed by Russo et al. [9]. Their approach was designed for Sentinel-1 land-cover mapping over geographically different regions including Amazonia, Africa, and Siberia. Rather than providing a dense sequence of individual SAR acquisitions directly to the network, the methodology organises acquisitions into temporal groups and derives a compact set of seasonal features. These features are stacked along the channel dimension and classified using a transformer-based U-shaped segmentation architecture.

The Philippine implementation preserves the main conceptual elements of that framework:

1. a set of seasonal Sentinel-1 VH representations is used instead of a raw dense time series;

2. the temporal and feature information is placed along the input-channel dimension;

3. the classification is formulated as dense semantic segmentation;
4. the scenes are subdivided into overlapping 256×256 training samples;
5. a Swin-transformer-based U-shaped architecture is used;
6. a batch size of one is used because of the relatively high-dimensional input and available GPU memory;
7. model performance is assessed from pixel-level comparison with a reference land-cover map.

At the same time, the Philippine experiment introduces several substantial adaptations. The Sentinel-1 features correspond to 2019 rather than the year used in the original experiments. The reference map is a Philippines-specific MOLCA-derived product developed in this thesis. The final legend contains ten classes, including a dedicated Mangrove Forest class. The valid training dataset consists of 277 scenes after data-availability screening. The training/validation split is constructed from the class composition of those scenes. Class-weighted loss is used to address the strongly imbalanced reference distribution, and a controlled weighted-sampling experiment is performed for Shrubland. Finally, the actual implementation employs the two-dimensional `SwinUNETR` model available in MONAI rather than reproducing a separate pure Swin-Unet code base.

Table 5.1 summarises the principal methodological relationship.

5.1.4 Overall deep learning workflow

The complete deep learning workflow begins only after the reference and SAR features described in Chapter 4 have been assembled. The principal stages are:

A central design principle is that operations capable of introducing statistical dependence between samples are performed *after* the scene-level split. In particular, overlapping sub-patches are not generated first and then randomly distributed between training and validation. Instead, the parent scenes are separated before training patches are constructed. This distinction is important because neighbouring overlapping patches can share up to a large proportion of their pixels.

The model is trained on the resulting training patches, while the validation scenes remain unchanged in membership throughout the controlled experiments. Consequently, modifications such as the weighted Shrubland sampler change only the exposure of training patches and not the underlying validation data.

Table 5.1: Relationship between the methodology of Russo et al. [9] and the Philippine implementation developed in this thesis. The table describes methodological differences rather than performance differences.

Aspect	**Russo et al. framework**	**Philippine implementation**
Study areas	Amazonia, Africa, and Siberia	Philippines
Reference data	MOLCA reference dataset	Philippines-specific MOLCA-derived reference, including the final revised reference used for the controlled DL experiments
SAR sensor	Sentinel-1	Sentinel-1
Polarization	VH	VH
Input organisation	Seasonally synthesised spatio-temporal features	Four temporal groups × seven feature representations = 28 channels
Segmentation family	Swin-Unet methodology	2-D MONAI SwinUNETR implementation following the same U-shaped Swin-transformer segmentation concept
Training image size	256 × 256 patches obtained from 512 × 512 scenes	256 × 256 overlapping patches obtained from 512 × 512 scenes
Output legend	Region-dependent MOLCA land-cover classes	Ten Philippine classes, including Mangrove Forest and No-Data
Class imbalance treatment	Original training configuration	Class-weighted cross entropy; additional controlled 10× Shrubland-patch sampling experiment
Model-selection strategy	Validation-based training	Validation mIoU with best-checkpoint saving and early stopping

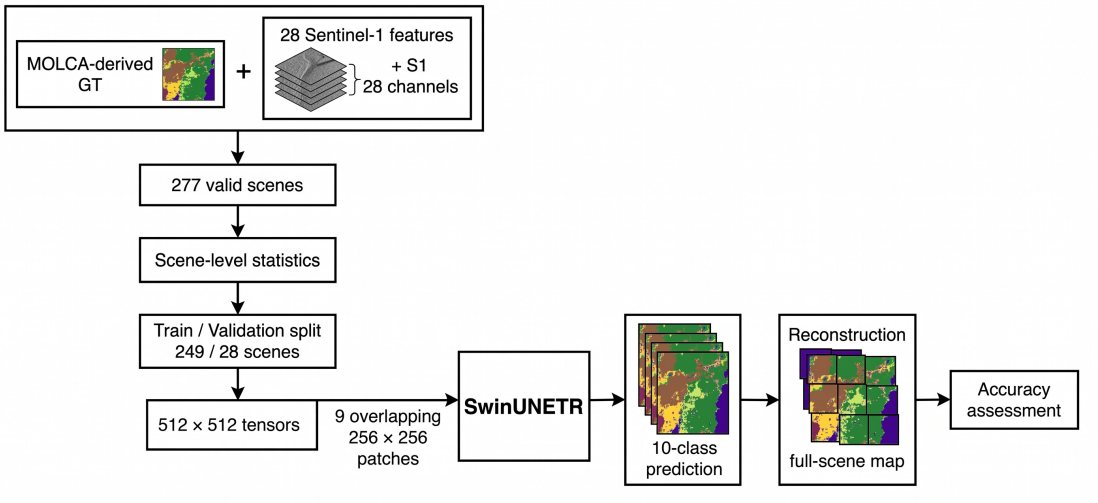


Figure 5.1: Overall deep learning workflow used for Sentinel-1 land-cover segmentation in the Philippines, from the MOLCA-derived reference scenes and 28-channel Sentinel-1 feature stacks to model training, validation, full-scene reconstruction, and accuracy assessment.

5.2 U-Net / Swin-UNETR Model Architecture

5.2.1 The U-shaped encoder–decoder principle

U-Net was introduced by Ronneberger et al. [13] as an encoder–decoder architecture for dense image segmentation. Its basic principle is particularly suitable for tasks in which both semantic context and precise localisation are required.

The contracting or encoder path progressively reduces the spatial resolution while increasing the richness of the extracted representations. At deeper levels of the network, individual features summarise progressively wider spatial contexts. The expanding or decoder path then reconstructs the original spatial resolution so that a class decision can be made for every pixel.

An essential component of the U-shaped design is the use of skip connections between corresponding encoder and decoder levels. Information from higher-resolution encoder layers is transferred directly to the decoder instead of forcing all spatial detail through the deepest bottleneck. This allows the output to combine high-level semantic information with fine-scale spatial structure.

For land-cover segmentation, the same principle has an intuitive interpretation. Large contextual structures such as a forest block, agricultural zone, river network, or urban area should influence the classification, but the final prediction must still preserve boundaries at the scale of individual pixels. A U-shaped architecture

provides a mechanism for combining these two requirements.

5.2.2 From convolutional U-Net to vision transformers

Traditional U-Net architectures are built primarily from convolutional operations. Convolutions have several advantages, including local connectivity, parameter sharing, and strong performance in image analysis. However, their direct receptive field is local. Wider spatial relationships are acquired only after multiple convolutional and downsampling operations.

Vision Transformers introduced an alternative way of modelling relationships between image elements by using self-attention. In a generic self-attention operation, an input representation $\mathbf{X}$ is projected into query, key, and value representations,

$$\mathbf{Q} = \mathbf{X}\mathbf{W}_Q, \qquad \mathbf{K} = \mathbf{X}\mathbf{W}_K, \qquad \mathbf{V} = \mathbf{X}\mathbf{W}_V, \tag{5.8}$$

and attention is calculated as

$$\text{Attention}\left(\mathbf{Q}, \mathbf{K}, \mathbf{V}\right) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V}, \tag{5.9}$$

where d is the dimensionality of the key representation.

The advantage of attention is that the response associated with one token can depend directly on other tokens. A straightforward global self-attention operation, however, becomes expensive for high-resolution images because the number of pairwise interactions increases rapidly with the number of image tokens.

5.2.3 Hierarchical Swin Transformer and shifted-window attention

The Swin Transformer addresses this computational limitation through a hierarchical representation and window-based self-attention [14]. Rather than computing attention globally between every image token at every layer, attention is restricted to local windows. This considerably reduces computational cost.

If windows remained permanently fixed, information exchange across their boundaries would be limited. The Swin architecture therefore alternates regular window partitioning with shifted windows. In one transformer block, self-attention is calculated inside one set of local windows; in a subsequent block, the window partition is shifted. Tokens that previously belonged to different windows can then interact.

The combination of windowed and shifted-window attention gives the architecture two properties that are important for dense land-cover mapping. First, it

preserves computational feasibility for high-resolution feature maps. Second, repeated hierarchical processing allows information to propagate beyond the initial local window and contributes to larger-scale contextual representation.

The Swin Transformer also constructs a hierarchy in which spatial resolution is progressively reduced, and feature dimensionality is increased. It can therefore serve as an encoder for U-shaped segmentation networks in a way that is conceptually comparable to the multi-resolution encoder of U-Net while replacing the principal contextual modelling mechanism with transformer blocks.

5.2.4 Swin-Unet methodological lineage

Cao et al. [35] extended the Swin concept into a U-shaped pure-transformer segmentation architecture. In Swin-Unet, tokenised image patches are processed by a hierarchical Swin Transformer encoder, while a symmetric decoder progressively restores the spatial resolution. Skip connections transfer multi-scale information between encoder and decoder stages.

Russo et al. [9] adopted this general Swin-Unet concept for Sentinel-1 land-cover mapping. Their work showed how the U-shaped transformer model can operate on the channel stack of seasonal SAR features rather than on conventional RGB images. The temporal-feature dimension is therefore incorporated into the network as input channels, while the network itself performs two-dimensional dense spatial segmentation.

The present thesis follows this methodological lineage but implements the network using MONAI's `SwinUNETR` class. This distinction is important for reproducibility. The terms "Swin-Unet" and "SwinUNETR" should not be treated as perfectly interchangeable architectural names.

The original Swin-Unet is a U-shaped pure-transformer segmentation model [35]. Hatamizadeh et al. [15] introduced Swin UNETR, which connects a hierarchical Swin Transformer encoder to a decoder through multi-scale skip connections. MONAI provides a maintained implementation of this model family [16].

For this reason, throughout the remainder of this chapter, *SwinUNETR* is used when referring specifically to the implemented Python model, whereas *Swin-Unet methodology* refers to the broader architectural lineage and to the terminology adopted by Russo et al. [9].

5.2.5 Two-dimensional SwinUNETR implementation

The model used in the Philippine experiments is instantiated using.

```
monai.networks.nets.SwinUNETR
```

with `spatial_dims=2`. Consequently, the network processes two-dimensional spatial samples while treating the 28 SAR representations as channels.

The network input for one training batch has the form.

$$\mathbf{X} \in \mathbb{R}^{1\times 28\times 256\times 256}, \tag{5.10}$$

because the training batch size is one. The network output is

$$\mathbf{Z} \in \mathbb{R}^{1\times 10\times 256\times 256}. \tag{5.11}$$

Each of the ten output channels contains the logits associated with one land-cover class.

The feature-size parameter is fixed to 48. This is consistent with the configuration investigated in the methodological work of Russo et al. [9] and provides a compromise between representational capacity and computational requirements.

The actual model constructor used for the experiments sets the conventional dropout probability, attention dropout probability, and stochastic depth/drop-path probability to zero. Gradient checkpointing is also disabled. The principal implemented parameters are summarised in Table 5.2.

Table 5.2: Principal SwinUNETR parameters used in the Philippine experiments.

Parameter	**Implemented value**
Model implementation	MONAI `SwinUNETR`
Spatial dimensionality	2-D
Input patch dimensions	256×256
Input channels	28
Output channels	10
Feature size	48
Standard dropout	0.0
Attention dropout	0.0
Drop-path rate	0.0
Gradient checkpointing	Disabled
Batch size	1
Primary model framework	PyTorch / MONAI

The code contains a compatibility mechanism for different MONAI versions. It first attempts to instantiate `SwinUNETR` with an explicit `img_size` argument. If the installed MONAI version does not accept that argument, the constructor is called again without it. This modification was introduced so that the project could remain executable across the software environments used during development. The model dimensions and input/output interpretation remain unchanged.

5.2.6 Encoder, bottleneck, decoder, and skip connections

At a conceptual level, the implemented network can be interpreted through the familiar encoder–bottleneck–decoder organisation.

The **encoder** transforms the 256×256 input through a hierarchy of Swin Transformer representations. At shallow levels, features retain relatively high spatial resolution. At deeper levels, the spatial representation becomes coarser while the feature representation becomes richer. Windowed and shifted-window attention allows the encoder to combine local information with progressively wider contextual relationships.

The **bottleneck** represents the deepest part of the network. At this level, the model contains a highly abstract representation of the input scene. Fine spatial detail is reduced, but contextual information is maximised.

The **decoder** progressively restores spatial resolution and transforms the encoded representation into a dense classification map. Skip connections transfer features from corresponding encoder stages so that spatial detail lost during downsampling can contribute directly during reconstruction.

Finally, a classification layer produces ten output scores for every spatial location.

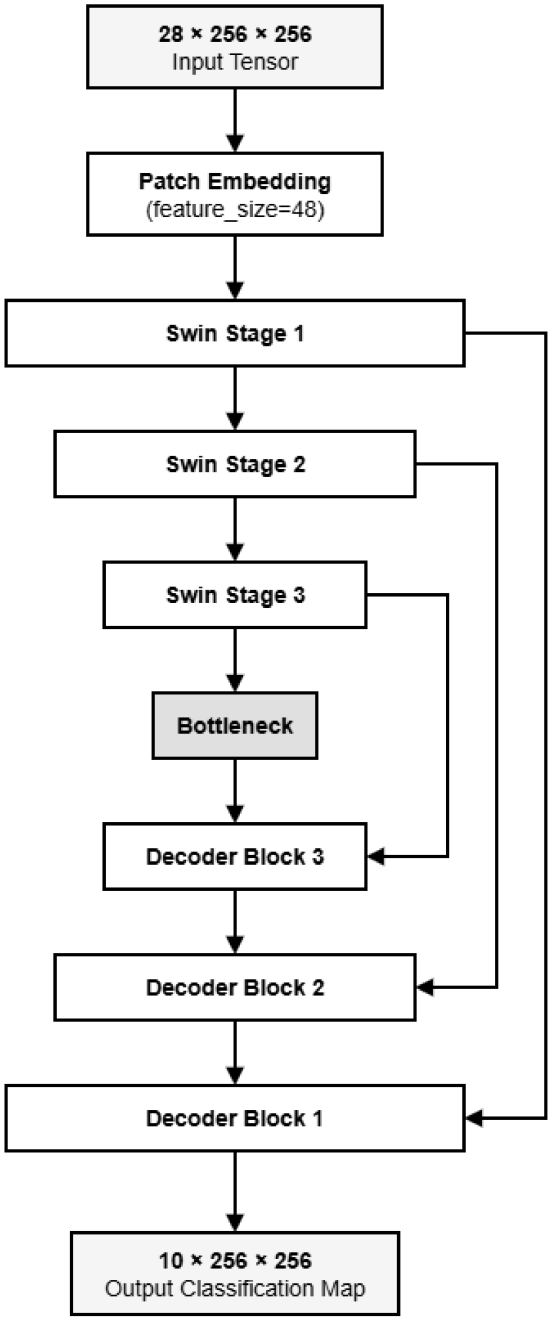


Figure 5.2: Simplified architecture of the 2-D MONAI SwinUNETR implementation used in this thesis.

5.2.7 Why a transformer-based U-shaped model is suitable for the present dataset

The selected architecture is well matched to the characteristics of the Philippine dataset for several reasons.

First, the input is multi-channel rather than RGB. The 28 channels encode spatial and temporal SAR information derived from four parts of the year. A deep encoder can learn combinations of these channels that are difficult to describe through manually chosen thresholds.

Second, many land-cover classes are spatially structured. Large forest regions, agricultural mosaics, rivers, coastlines, wetland complexes, mangrove belts, and built environments exhibit different contextual patterns. Hierarchical attention provides a mechanism for representing relationships beyond the immediate pixel neighbourhood.

Third, dense classification requires accurate spatial reconstruction. The U-shaped decoder and skip connections are specifically designed for this purpose.

Fourth, the training dataset is much smaller than datasets commonly used for generic computer-vision pre-training. The use of overlapping samples, data augmentation, class weighting, validation-based checkpoint selection, and early stopping therefore becomes an integral part of the methodology rather than a secondary implementation detail.

5.3 Input Features and Output Classes

5.3.1 Twenty-eight-channel Sentinel-1 representation

The input to the segmentation model is the Sentinel-1 representation generated in Chapter 4. Only the VH polarisation is used. The 2019 acquisitions are organised into four temporal groups, and seven representations are produced for each group:

$$4 \text{ temporal groups} \times 7 \text{ representations} = 28 \text{ input channels.} \tag{5.12}$$

The seven representations are:

1. seasonal Super Image (SI);
2. Lee-filtered Super Image;
3. maximum (MAX);
4. minimum (MIN);
5. maximum–minimum difference (MAXMIN);

6. arithmetic mean (MEAN);

7. median (MEDIAN).

Their physical construction was discussed in Chapter 4. The present chapter focuses on how these rasters are organised for neural-network processing.

The feature ordering must remain fixed because the network associates each input channel with a learned set of parameters. Interchanging channels between training and inference would therefore change the meaning of the model input. In the project pipeline, each reference scene has an associated JSON file defining the ordered list of its 28 Sentinel-1 feature rasters. The JSON association, rather than an uncontrolled filename sort, is used as the source of truth for feature ordering.

Table 5.3 shows the conceptual organisation.

Table 5.3: Conceptual organization of the 28 Sentinel-1 input channels.

Temporal group	**Seven feature representations**
Group 1	SI, LEE, MAX, MIN, MAXMIN, MEAN, MEDIAN
Group 2	SI, LEE, MAX, MIN, MAXMIN, MEAN, MEDIAN
Group 3	SI, LEE, MAX, MIN, MAXMIN, MEAN, MEDIAN
Group 4	SI, LEE, MAX, MIN, MAXMIN, MEAN, MEDIAN
Total	28 channels

5.3.2 Scene–feature association and integrity checking

Every valid deep learning sample is defined by three elements:

$$S_i = (G_i, \{F_{i,1}, F_{i,2}, \ldots, F_{i,28}\}, J_i)\,, \tag{5.13}$$

where G_i denotes the reference raster, $F_{i,t}$ denotes the t-th SAR feature raster, and J_i denotes the metadata/association JSON.

Before training, the dataset-indexing stage verifies that the required files exist and that the expected feature association can be constructed. This is important because a missing feature would otherwise alter the number or ordering of channels and make the sample incompatible with the trained network.

The invalid Sentinel-2 tile case described in Chapter 4 was detected during this preparation stage. The original selection contained 280 candidate samples, but three samples associated with tile `51QUC` could not be retained because the required valid 2019 VH Sentinel-1 feature set was not available. The final deep learning dataset therefore contains

$$N_{\text{scenes}} = 277 \tag{5.14}$$

valid scene–reference associations.

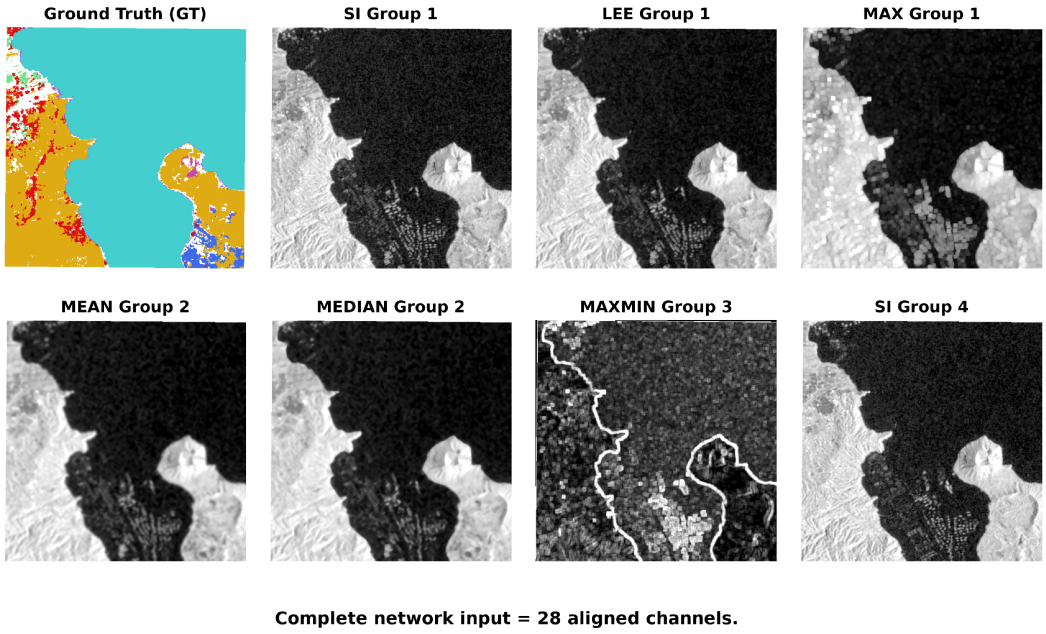


Figure 5.3: Representative 28-channel Sentinel-1 input stack alongside the MOLCA-derived ground truth for scene 51PTR_09_15. Seven selected channels from different temporal groups are shown to illustrate the varying SAR representations. The complete network input consists of 28 aligned channels, whose exact sequence is defined by the scene's associated JSON file.

5.3.3 Spatial resizing to the network scene size

The geographic samples selected from the Sentinel-2 framework have the original patch dimensions described in Chapter 4. For the deep learning pipeline, the reference raster and each Sentinel-1 feature raster are resized to

$$512 \times 512 \tag{5.15}$$

pixels before sub-patching.

The current implementation uses nearest-neighbour interpolation for this resizing operation. For the categorical reference raster, nearest-neighbour interpolation is required because interpolating between numerical class codes would generate invalid semantic categories.

The current loader applies the same nearest-neighbour resizing operation to the Sentinel-1 feature rasters. This is stated explicitly here for reproducibility: it describes the implementation used for the reported experiments rather than a generic recommendation concerning continuous SAR-image resampling.

Any NaN values encountered during loading are converted to zero after resizing. The reference values are also remapped from their stored raster codes to the internal class IDs used by the neural network.

5.3.4 Sentinel-1 normalization

Before being supplied to the network, the Sentinel-1 feature rasters are normalised by the finalised dataset-loading pipeline. The same preprocessing configuration is used for training, validation, and inference so that each feature occupies the same numerical convention whenever it is presented to the model.

The final input representation contains all seven feature families—SI, LEE, MAX, MIN, MAXMIN, MEAN, and MEDIAN—for each of the four temporal groups. Normalisation is applied according to the configured rule for the corresponding feature representation and the resulting values are converted to the numerical range expected by the network. The channel order is preserved throughout this operation, so normalisation changes the numerical scale of a feature but never its position within the 28-channel tensor.

Any non-finite values remaining after spatial preparation are converted to zero before tensor construction. The categorical reference raster is processed separately and remapped only to the integer class identifiers defined by the Philippine legend.

Using one finalised preprocessing path for all controlled experiments is important because the clean CC50 baseline and the weighted-sampling experiment are intended to differ in training-patch sampling, not in their input data or normalisation procedure.

5.3.5 Tensor organization

After loading, normalisation, resizing, and stacking, one scene is represented as

$$\mathbf{X}_i^{(512)} \in \mathbb{R}^{28\times512\times512}. \tag{5.16}$$

Its corresponding reference is

$$\mathbf{Y}_i^{(512)} \in \{0,\dots,9\}^{512\times512}. \tag{5.17}$$

The temporal-feature axis is therefore treated as the channel axis. The network does not receive an explicit three-dimensional tensor with a separate temporal axis.

Instead, temporal and derived-feature information is encoded by channel position, following the strategy used by Russo et al. [9].

This representation can be written as

$$\mathbf{X}_i = [\mathbf{F}_{i,1}, \mathbf{F}_{i,2}, \ldots, \mathbf{F}_{i,28}] \,, \tag{5.18}$$

where every $\mathbf{F}_{i,t}$ is a spatial raster aligned with the same reference scene.

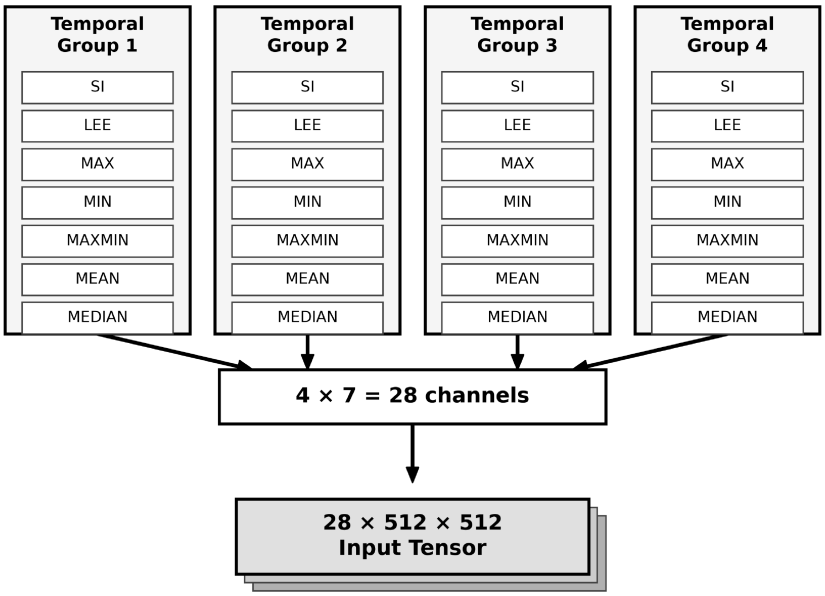


Figure 5.4: Conceptual structure of the input tensor. The 28 Sentinel-1 channels are derived from seven feature representations calculated across four temporal groups, which are subsequently stacked into a single $28 \times 512 \times 512$ input tensor prior to sub-patching.

5.3.6 Output land-cover legend

The output layer contains ten channels corresponding to the class system developed for the Philippine reference map. The internal neural-network IDs are listed in Table 5.4.

The distinction between NoData and Mangrove Forest is important. Earlier versions of the underlying software were developed for different regional legends. In the Philippine adaptation, class ID 9 is a valid Mangrove Forest category and

Table 5.4: Output classes used by the deep learning model.

ID	Class	Role during optimization
0	NoData / unclassified	Ignored as reference target
1	Built-up	Trained class
2	Cropland	Trained class
3	Forest	Trained class
4	Grassland	Trained class
5	Shrubland	Trained class
6	Wetland	Trained class
7	Water	Trained class
8	Bareland	Trained class
9	Mangrove Forest	Trained class

must therefore remain available to the model. Only class ID 0 is excluded from the target pixels used to calculate the loss.

The model still contains ten output channels, including the NoData channel, so that the internal output coding remains aligned with the reference legend. However, pixels for which the ground-truth target is class 0 do not contribute to the optimisation objective or the principal evaluation metrics.

If the model predicts class 0 at a location where the reference pixel belongs to a valid land-cover class, that prediction is not ignored. It is counted as an incorrect prediction. The masking is based on the *reference target*, not simply on the predicted class.

5.3.7 Overlapping training patches

The 512×512 tensors are not passed directly through the model during training. Each scene is divided into overlapping 256×256 samples using a stride of 128 pixels.

For one spatial dimension, the number of extracted patches is

$$n = \left\lfloor \frac{H - P}{S} \right\rfloor + 1, \tag{5.19}$$

where $H = 512$ is the scene dimension, $P = 256$ is the patch size, and $S = 128$ is the stride. Therefore,

$$n = \left\lfloor \frac{512 - 256}{128} \right\rfloor + 1 = 3. \tag{5.20}$$

The two-dimensional number of patches is

$$N_p = 3 \times 3 = 9. \tag{5.21}$$

The top-left patch coordinates are therefore generated from the spatial positions.

$$\{0, 128, 256\} \tag{5.22}$$

in both dimensions.

The 50% overlap creates multiple contextual views of the central regions of each scene. It also increases the number of training samples without requiring the creation of synthetic SAR observations. This follows the patching strategy used in the methodological framework of Russo et al. [9].

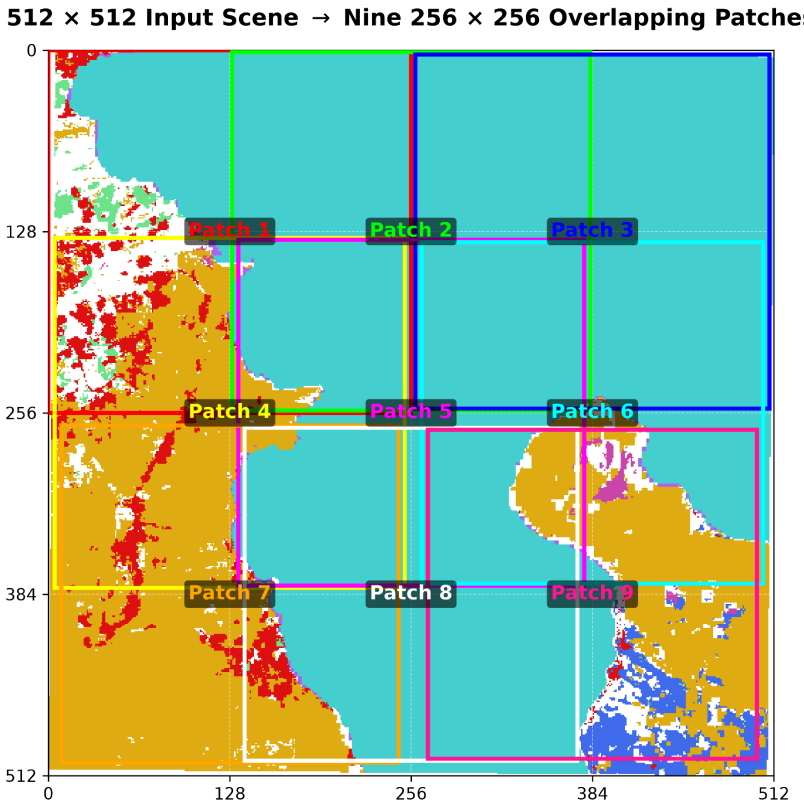


Figure 5.5: Patching strategy used to transform each 512 × 512 scene into nine overlapping 256 × 256 network inputs. The 50% overlap (stride of 128 pixels) ensures that the network receives multiple contextual views of the central regions of each scene.

5.4 Training, Validation, and Testing Strategy

5.4.1 Final deep learning sample population

The deep learning pipeline begins with the 277 valid scene associations. Each association contains a reference raster and all 28 Sentinel-1 feature rasters required by the model.

The separation into training and validation sets is performed using the parent 512×512 scenes. This ordering is deliberate:

$$277 \text{ scenes} \rightarrow \text{scene-level split} \rightarrow \text{overlapping sub-patches}. \tag{5.23}$$

The alternative order,

$$\text{scenes} \rightarrow \text{overlapping patches} \rightarrow \text{random patch split}, \tag{5.24}$$

was avoided because patches from the same parent scene share both geographic context and pixels. If such patches were allowed in both training and validation, the validation set would not provide a sufficiently independent assessment of the model's ability to classify unseen scenes.

5.4.2 Scene-level class-composition vectors

A purely random scene split can be problematic for an imbalanced multi-class dataset. If the validation subset were accidentally dominated by only the common classes, or if a specific landscape composition occurred almost exclusively in one partition, the resulting model comparison would be less informative.

To reduce this risk, a class-composition vector is calculated for every scene before splitting. After reference-code remapping and exclusion of the ignored class, the proportion of each class is computed as

$$p_{i,c} = \frac{n_{i,c}}{\sum_{j \in \mathcal{C}_v} n_{i,j}}, \tag{5.25}$$

where $n_{i,c}$ is the number of valid pixels of class c in scene i and $\mathcal{C}_v$ denotes the valid land-cover classes.

The resulting scene descriptor is

$$\mathbf{p}_i = [p_{i,0}, p_{i,1}, \ldots, p_{i,9}]\,. \tag{5.26}$$

Because NoData pixels are excluded from the valid-pixel denominator, the valid class proportions characterise the land-cover composition of the usable area of each scene.

5.4.3 K-means grouping and stratified scene split

The 277 scene-composition vectors are grouped with K-means clustering using.

$$K_{\text{clusters}} = 4. \tag{5.27}$$

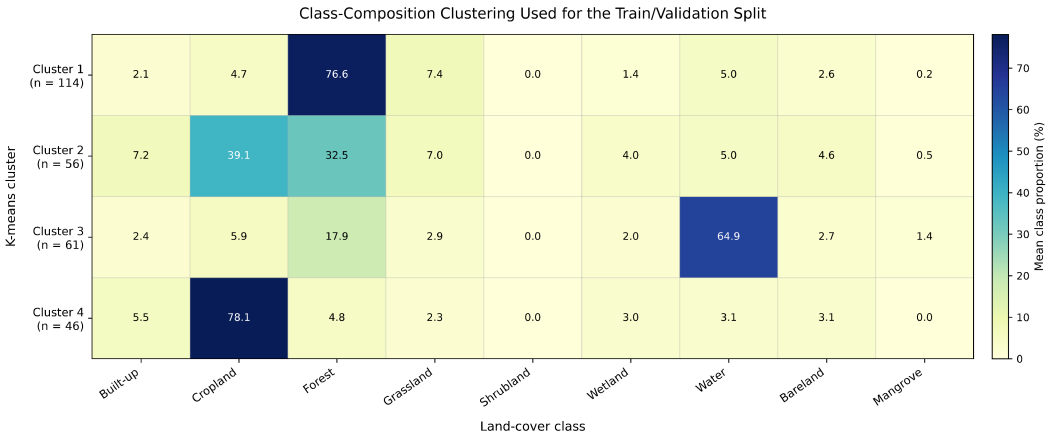


Figure 5.6: Mean land-cover class proportions for the four K-means clusters used to stratify the scene split. By grouping scenes according to their ten-dimensional class-composition vectors, the split ensures that different dominant landscape types (such as predominantly Forest, Water, or Cropland scenes) are proportionally represented in both the training and validation partitions, avoiding the biases of arbitrary random sampling.

The clustering random state is 42, and the implementation uses ten K-means initialisations. The purpose is not to assign ecological meaning to the resulting clusters. Instead, the clusters provide coarse strata representing different types of class composition.

A stratified shuffle split is subsequently performed using these cluster labels. The validation fraction is

$$f_{\text{val}} = 0.10. \tag{5.28}$$

This produces

$$N_{\text{train}} = 249 \tag{5.29}$$

training scenes and

$$N_{\text{val}} = 28 \tag{5.30}$$

validation scenes.

The same split is retained for the controlled final experiments. Reusing the same validation membership is important when comparing modifications to the training procedure because otherwise a change in performance could partly be caused by a different validation sample rather than by the methodological intervention itself.

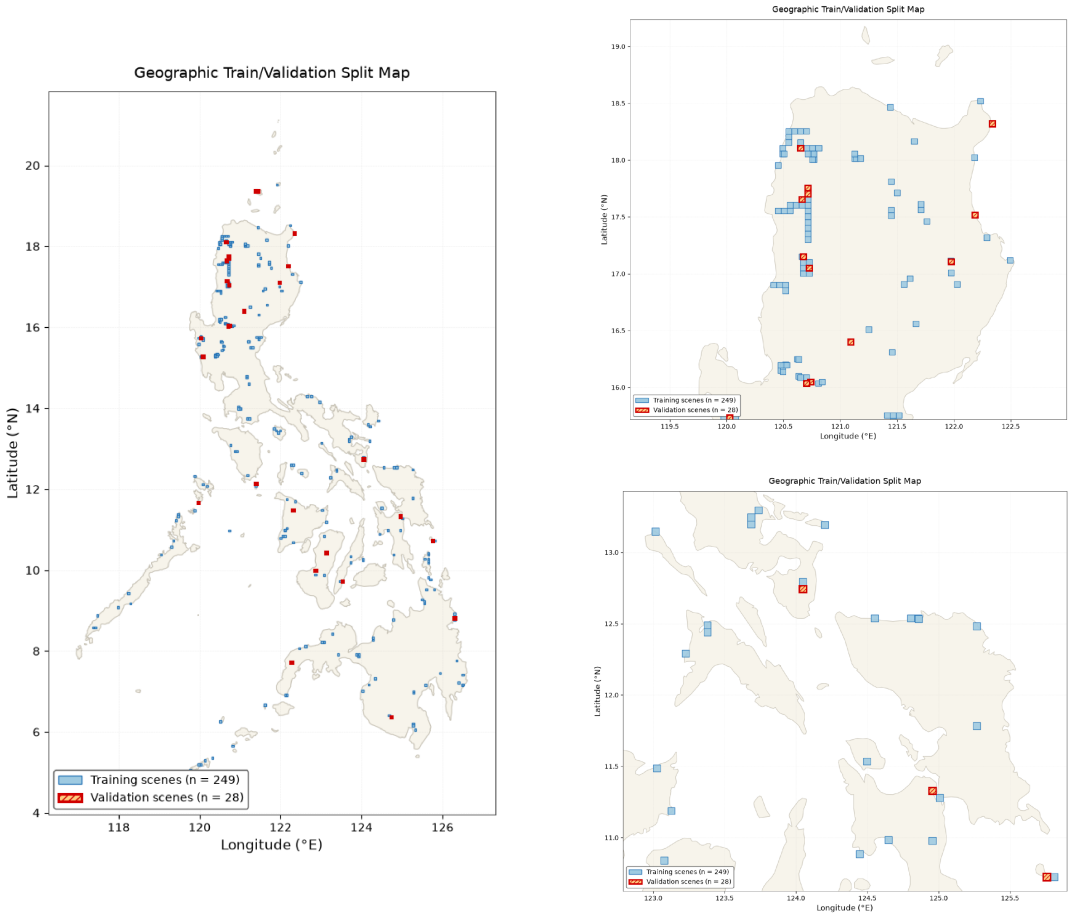


Figure 5.7: Geographic distribution of the selected Sentinel-1 scenes across the Philippines. The dataset is explicitly partitioned into 249 training scenes (light blue squares) and 28 validation scenes (red hashed squares) utilising the `tile_split.json` configuration. This mapping visualises the geographic outcome of the class-composition stratified split, demonstrating that both data subsets successfully maintain a representative and intermixed spatial distribution across the archipelagic extent.

5.4.4 Number of training and validation patches

Because every scene generates nine 256×256 patches, the training set contains

$$249 \times 9 = 2241 \tag{5.31}$$

training patches, while the validation set contains

$$28 \times 9 = 252 \tag{5.32}$$

validation patches.
Table 5.5 summarises the hierarchy.

Table 5.5: Scene and patch counts used in the deep learning pipeline.

Quantity	**Value**
Originally selected reference patches	280
Samples excluded because of invalid/missing required S1 inputs	3
Valid deep learning scenes	277
Training scenes	249
Validation scenes	28
Sub-patches generated per scene	9
Training sub-patches	2241
Validation sub-patches	252
Network patch dimensions	256×256
Spatial stride	128 pixels
Input channels per sub-patch	28
Output classes	10

5.4.5 Training-only data augmentation

The training loader applies two simple geometric augmentations:

$$P(\text{horizontal flip}) = 0.5 \tag{5.33}$$

and

$$P(\text{vertical flip}) = 0.5. \tag{5.34}$$

The two decisions are sampled independently.

The same geometric operation is applied to the input tensor and the corresponding reference mask, so pixel alignment between SAR features and target labels is preserved. No augmentation is applied to validation data.

The use of flips is suitable for the present task because the semantic identity of a land-cover type does not depend on image orientation. A forest remains Forest after horizontal or vertical reflection, as do the other target classes. The operation can therefore increase the range of spatial configurations observed during training without changing class semantics.

Rotation, spectral perturbation, synthetic noise injection, or more aggressive geometric transformations were not part of the principal training configuration described here.

5.4.6 In-memory patch construction and data loading

After the scene-level datasets have been created, their overlapping patches are constructed in memory. The resulting training and validation tensors are wrapped in PyTorch `TensorDataset` objects.

The clean baseline uses ordinary randomised training order:

$$\text{shuffle} = \text{True} \tag{5.35}$$

for the training DataLoader and

$$\text{shuffle} = \text{False} \tag{5.36}$$

for validation.

The batch size is one,

$$B = 1, \tag{5.37}$$

and the number of DataLoader workers is set to zero. This conservative configuration was selected for the available workstation and avoids multiprocessing-related variability while keeping GPU memory requirements manageable.

Because each batch contains a 28-channel 256×256 input, the model processes

$$28 \times 256 \times 256 = 1\,835\,008 \tag{5.38}$$

Input values per training sample before any internal feature expansion by the network.

5.4.7 Reproducibility settings

A seed value of 42 is used in the principal configuration. It controls the train–validation splitting procedure and the PyTorch random state used during model initialisation and sampling. CUDA random seeds are also initialised when a CUDA device is available.

It should be noted that exact bitwise reproducibility across different software versions, GPU architectures, and all stochastic preprocessing operations is not claimed. In particular, geometric augmentation in the dataset loader uses the Python random-number generator. The purpose of the seed configuration is therefore to stabilise the main experimental procedure and fixed split rather than to claim that every low-level operation is necessarily deterministic across all hardware and software environments.

5.4.8 Clean CC50 baseline experiment

Following the reference-map investigation described in Chapter 3, the CC50 version was adopted as the principal revised reference dataset for the final controlled deep learning experiments.

The *CC50 clean baseline* is defined as the training protocol in which:

$$\text{training patches are sampled normally} \tag{5.39}$$

without a class-specific patch sampler.

The term "clean" does not mean that no class-imbalance correction is applied. The clean baseline still uses the class-weighted cross-entropy described in Section 5.5. The distinction is that the frequency with which complete training patches are presented is not modified according to their Shrubland content.

This baseline establishes a controlled reference against which the sampling experiment can be compared.

5.4.9 Shrubland-focused WeightedRandomSampler experiment

A second controlled experiment modifies only the sampling of training patches. Its purpose is to test whether increasing exposure to patches containing the rare Shrubland class can improve the model's opportunity to learn Shrubland-related features.

Let

$$s_i = \begin{cases} 1, & \text{if training patch } i \text{ contains at least one Shrubland pixel,} \\ 0, & \text{otherwise.} \end{cases} \tag{5.40}$$

The sampling weight assigned to patch i is

$$q_i = \begin{cases} 10, & s_i = 1, \\ 1, & s_i = 0. \end{cases} \tag{5.41}$$

The criterion is therefore binary at the patch level. A patch receives the 10× sampling weight whenever *at least one* pixel has Shrubland label 5. The amount of Shrubland inside that patch does not further scale its sampling weight.

Within the 2241 training patches,

$$N_{\text{shrub patch}} = 296 \tag{5.42}$$

and

$$N_{\text{nonshrub patch}} = 1945. \tag{5.43}$$

Without weighted sampling, the Shrubland-containing patch fraction is therefore

$$f_{\text{original}} = \frac{296}{2241} \approx 0.1321. \tag{5.44}$$

With weights 10 and 1, the expected probability of drawing a Shrubland-containing patch becomes

$$f_{\text{sampled}} = \frac{10 \times 296}{10 \times 296 + 1 \times 1945} \approx 0.6035. \tag{5.45}$$

Sampling is performed *with replacement*, and the number of draws per epoch remains 2241. Consequently, the number of optimisation steps per epoch is not increased; instead, the composition of the patches presented during those steps is altered.

This distinction is important. The sampling experiment does not create new reference labels and does not synthetically generate new Shrubland pixels. It repeatedly exposes the model to existing patches in which Shrubland occurs.

The validation DataLoader is unchanged. The scene split, validation scenes, input features, model architecture, loss class weights, optimiser, scheduler, and evaluation procedure are also preserved. The experimental comparison therefore isolates patch-level training exposure as far as practical.

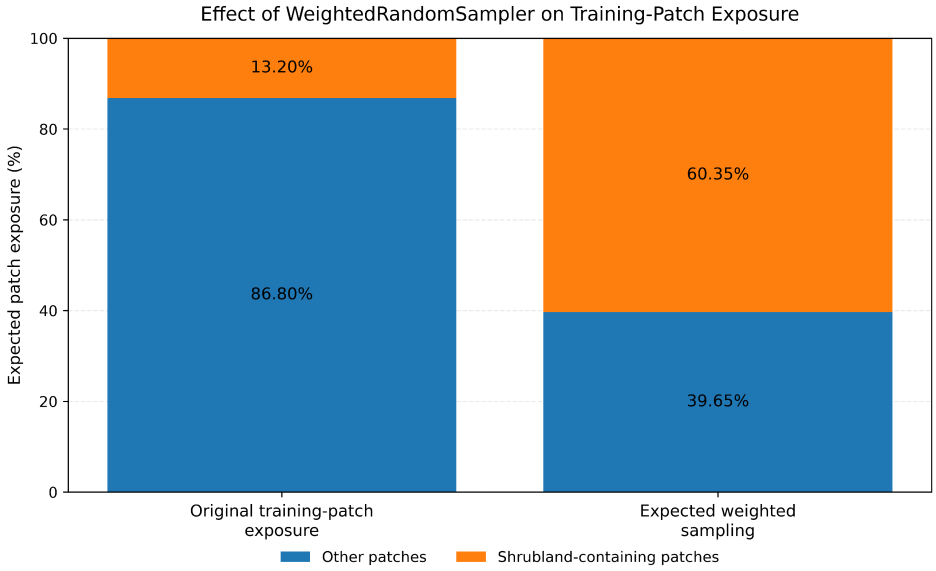


Figure 5.8: Effect of the WeightedRandomSampler on training-patch exposure during the Shrubland-focused experiment. By assigning a sampling weight of 10 to any patch containing at least one Shrubland pixel and sampling with replacement, the expected probability of exposing the model to Shrubland during a training step increases from 13.20% to 60.35%. This isolates patch-level exposure without synthetically generating new reference pixels.

5.4.10 Two levels of imbalance handling

The weighted sampler should not be confused with class weighting in the loss function.

Class weighting acts at the **pixel-loss level**. If a Shrubland pixel appears in a batch, its classification error contributes more strongly to the loss than an error for a frequent class.

The WeightedRandomSampler acts at the **patch-exposure level**. It makes a complete image patch containing Shrubland more likely to be selected for an optimisation step.

Conceptually,

$$\boxed{\text{class weighting} = \text{how strongly a pixel error matters}} \tag{5.46}$$

while

$$\boxed{\text{weighted sampling} = \text{how often a type of patch is seen}}. \tag{5.47}$$

Both approaches are established responses to imbalanced learning problems, in which strongly under-represented classes may otherwise have limited influence on empirical-risk minimization [36]. However, increasing the exposure of a minority class cannot by itself guarantee class separability. If the available reference labels are noisy, extremely sparse, geographically concentrated, or not sufficiently separable in feature space, resampling may have limited effect. This distinction becomes important when interpreting the experimental findings in Chapter 6.

5.4.11 Validation-based model selection and early stopping

The maximum configured training duration for the final controlled experiments is

$$E_{\max} = 100 \text{ epochs}. \tag{5.48}$$

Training is not required to reach epoch 100. The model-selection variable is validation mean Intersection over Union (mIoU). The best model is defined by improvement in this monitored metric.

For every epoch, the training loop performs:

$$\text{training} \rightarrow \text{validation} \rightarrow \text{monitor check} \rightarrow \text{checkpoint / patience update}. \tag{5.49}$$

A warm-up period of five epochs is used before early stopping can terminate the run. The minimum improvement threshold is

$$\Delta_{\min} = 0.001. \tag{5.50}$$

For the final CC50 controlled experiments, the early-stopping patience is 25 epochs. Thus, once the warm-up period has passed, training terminates if validation mIoU fails to improve by more than the minimum threshold for 25 consecutive epochs.

When an improvement occurs, the current model state is saved as

`best_model.pt`

together with the epoch, optimiser state, monitored metric, best metric value, and training configuration.

This strategy serves two purposes. It prevents unnecessary continuation once validation performance no longer improves and, more importantly, ensures that final evaluation uses the checkpoint associated with the best monitored validation state rather than automatically using the final chronological epoch.

5.4.12 Evolution of the early-stopping protocol

The training procedure evolved during the diagnostic stages of the project. The first complete Philippine baseline used a maximum of 100 epochs with a five-epoch warm-up, a minimum improvement of 0.001, and a patience of 15 epochs.

For the later controlled CC50 experiments, the patience was increased to 25 while the same general validation-mIoU monitoring strategy was retained. This distinction is maintained in the thesis because the early diagnostic model and the final controlled CC50 experiments should not be presented as if they were generated under an identical stopping configuration.

The consequences of the different runs are reported chronologically in Chapter 6. For the final methodology described here, the principal CC50 protocol uses the settings summarised in Table 5.8.

5.4.13 Absence of an independent held-out test partition

The implemented Philippine pipeline contains a training partition and a scene-level validation partition. A third independent test partition was not created from the 277 available scenes.

The validation partition therefore has two related functions:

1. It is used during model development for validation-mIoU monitoring, early stopping, and checkpoint selection;

2. the selected checkpoint is subsequently evaluated in greater detail on the same fixed validation scenes using full-scene reconstruction and class-wise metrics.

For this reason, the resulting accuracy values should be interpreted as *internal validation performance* rather than as an estimate obtained from a completely untouched external test dataset.

This wording is maintained throughout the thesis to avoid overstating the independence of the reported assessment. An independent geographic test dataset, additional expert-labelled samples, or evaluation on separately selected Philippine regions would provide a stronger external generalisation assessment and is therefore discussed as a direction for future work.

5.4.14 Computational environment

The principal full-training experiments were executed on the laboratory Linux workstation after the complete pipeline had first been checked through shorter diagnostic runs.

The principal environment used for the established training pipeline is summarised in Table 5.6.

Table 5.6: Principal computational environment used for full model training.

Component	**Configuration**
Operating system	Ubuntu 18.04.6 LTS
CPU	Intel Core i7-7700 @ 3.60 GHz
System memory	Approximately 62 GB RAM
GPU	NVIDIA GeForce GTX 1080 Ti
GPU memory	Approximately 11 GB
CUDA environment	CUDA 11.3 compatible environment
Python	Python 3.8
PyTorch	1.12.1 + CUDA 11.3 build
MONAI	1.2.0
Primary execution device	CUDA GPU

The environment was deliberately isolated in a project-specific Conda environment so that existing laboratory environments were not modified. Compatibility adjustments were made where necessary, including the model-constructor handling described in Section 5.2.5.

5.5 Loss Function and Optimisation

5.5.1 Class imbalance in the training reference

The Philippine reference dataset is strongly imbalanced. Large and widespread classes occupy considerably more pixels than some of the less frequent categories.

A standard unweighted loss assigns equal cost to every individual pixel. Under severe imbalance, this means that the total loss can be dominated by errors from the common classes simply because those classes contribute many more training pixels.

This problem is well established in imbalanced learning [36]. In a dense land-cover context, imbalance occurs at multiple levels: total pixel frequency, number of scenes containing a class, spatial extent within individual scenes, and geographic diversity of examples.

The present pipeline addresses the pixel-frequency component using class-weighted cross-entropy. The later controlled sampler experiment additionally modifies patch exposure, as described in Section 5.4.9.

5.5.2 Valid-pixel masking

NoData pixels are not meaningful supervised land-cover targets. They are therefore excluded from the optimisation objective.

Let the set of valid target pixels in a batch be

$$\Omega_v = \{i \mid y_i \neq 0\}\,. \tag{5.51}$$

The training loop explicitly constructs this validity mask and removes ignored pixels before computing the loss. If a patch contains no valid target pixels, that batch does not contribute an optimisation update.

The same ground-truth-driven masking principle is used for the principal accuracy metrics. This prevents the potentially large NoData area from inflating apparent model performance.

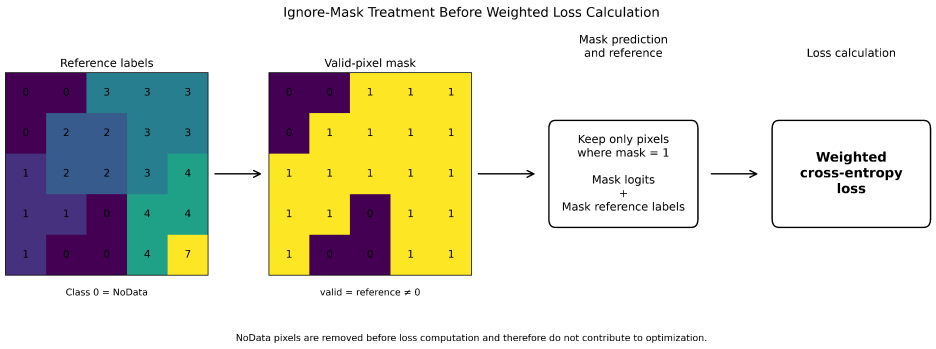


Figure 5.9: Illustration of the valid-pixel masking operation prior to loss calculation. Pixels corresponding to the NoData/unclassified category (class ID 0) are excluded from the target mask, ensuring that unlabelled regions do not contribute to the weighted cross-entropy optimisation.

5.5.3 Weighted categorical cross-entropy

For a valid target pixel i belonging to class y_i, the weighted cross-entropy contribution is

$$\ell_i = -w_{y_i} \log p_{i,y_i}, \tag{5.52}$$

where w_{y_i} is the class-specific weight and p_{i,y_i} is the predicted probability of the correct class.

For N_v valid pixels, the conceptual loss can be expressed as

$$\mathcal{L}_{\mathrm{WCE}} = -\frac{1}{N_v} \sum_{i \in \Omega_v} w_{y_i} \log p_{i,y_i}. \tag{5.53}$$

In the implementation, PyTorch's `CrossEntropyLoss` receives the raw model logits and performs the log-softmax/cross-entropy calculation internally. The NoData class is configured as the ignored target.

5.5.4 Computation of class-frequency weights

The class weights are derived only from the *training scenes*. Validation information is not used in their estimation.

An important implementation detail is that the weighting frequency is based on the average of scene-level class proportions rather than simply pooling every valid training pixel into one national histogram.

Let

$$p_{i,c} \tag{5.54}$$

be the valid-pixel proportion of class c in training scene i. For $N_t = 249$ training scenes, the estimated training frequency is

$$\bar{p}_c = \frac{1}{N_t} \sum_{i=1}^{N_t} p_{i,c}. \tag{5.55}$$

An unnormalized inverse-square-root weight is calculated as

$$\tilde{w}_c = \frac{1}{\sqrt{\bar{p}_c + \epsilon}}, \tag{5.56}$$

with

$$\epsilon = 10^{-6}. \tag{5.57}$$

The NoData weight is explicitly set to zero.

For the set of present non-ignored classes $\mathcal{C}_p$, the weights are normalized by their mean,

$$w_c = \frac{\tilde{w}_c}{\frac{1}{|\mathcal{C}_p|}\sum_{j\in\mathcal{C}_p}\tilde{w}_j}. \tag{5.58}$$

This normalisation keeps the average weight of the present trainable classes approximately equal to one while preserving their relative importance.

5.5.5 Final CC50 class weights

The class weights generated for the final CC50 training split are shown in Table 5.7.

Table 5.7: Inverse-square-root class weights used for the final CC50 training experiments. The weights were computed exclusively from the 249 training scenes.

ID	Class	Weight
0	NoData	0.0000
1	Built-up	0.6563
2	Cropland	0.2637
3	Forest	0.1965
4	Grassland	0.5624
5	Shrubland	3.5999
6	Wetland	0.8742
7	Water	0.3042
8	Bareland	0.7373
9	Mangrove Forest	1.8056

The relative magnitude of the weights follows the training distribution. Frequent classes such as Forest receive comparatively small weights, whereas underrepresented classes receive larger weights. Shrubland has the largest weight in the final CC50 configuration, while Mangrove Forest receives the second-largest value.

The class weights remain the same in the clean CC50 and 10× sampler experiments. Therefore, the sampler experiment does not replace weighted cross-entropy; it adds an additional patch-level exposure mechanism while preserving the pixel-level loss weighting.

5.5.6 Adam optimization

Network parameters are optimised with Adam [37]. Adam combines adaptive per-parameter scaling with estimates of the first and second moments of the gradient.

Let g_t denote the gradient at optimisation step t. The first-moment estimate is

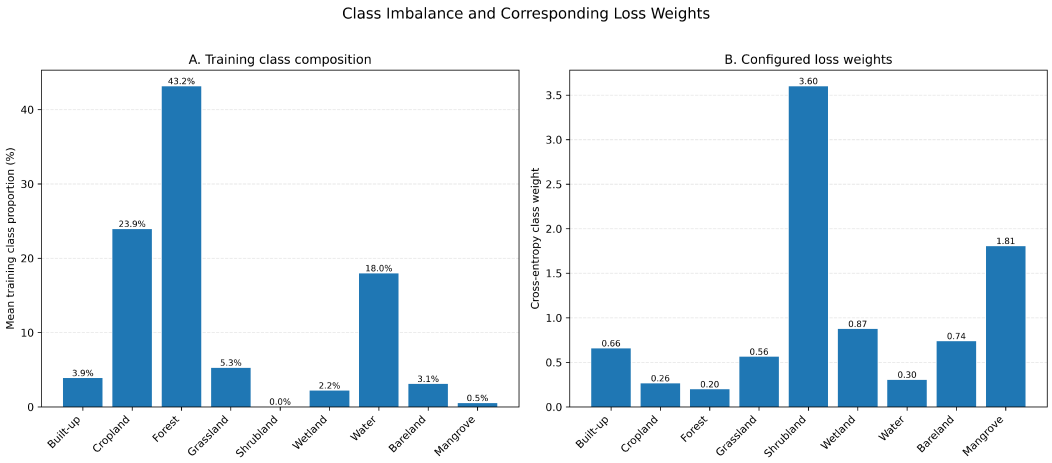


Figure 5.10: Comparison of the mean training class composition (Panel A) and the resulting inverse-square-root loss weights (Panel B). The visualisation demonstrates how dominant classes like Forest receive minimal weight, while extremely rare classes like Shrubland and Mangrove Forest are heavily penalised for misclassification to counter the severe pixel-level imbalance.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \tag{5.59}$$

and the second-moment estimate is

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2. \tag{5.60}$$

The bias-corrected estimates are

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \tag{5.61}$$

and

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}. \tag{5.62}$$

The generic Adam update is then.

$$\theta_t = \theta_{t-1} - \eta_t \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon_{\text{Adam}}}, \tag{5.63}$$

with the optimiser's standard numerical stabilisation term.
The initial learning rate used in this thesis is

$$\eta_0 = 10^{-4}. \tag{5.64}$$

This value is also consistent with the Swin-Unet configuration investigated by Russo et al. [9].

5.5.7 Weight decay

The Adam optimiser is configured with

$$\lambda_{\mathrm{wd}} = 0.01 \tag{5.65}$$

through its `weight_decay` parameter.

This should not be confused with using the separate AdamW optimiser. The implemented optimiser is `torch.optim.Adam`. Consequently, the description in this thesis follows the optimiser that actually ran rather than referring generically to decoupled AdamW regularisation.

5.5.8 Cosine learning-rate annealing

The learning rate is adjusted during training using PyTorch's `CosineAnnealingLR`. The configured parameters are

$$T_{\max} = 100 \tag{5.66}$$

and

$$\eta_{\min} = 10^{-6}. \tag{5.67}$$

A convenient form of the cosine schedule is

$$\eta(t) = \eta_{\min} + \frac{1}{2}\left(\eta_{\max} - \eta_{\min}\right)\left[1 + \cos\left(\frac{\pi t}{T_{\max}}\right)\right]. \tag{5.68}$$

Cosine learning-rate schedules were popularised in the SGDR family of methods [38]. In the present implementation, however, no periodic warm restart is performed. A single cosine-annealing scheduler is used, and training may terminate before $T_{\max}$ if the early-stopping condition is reached.

The scheduler is advanced after each completed training/validation epoch provided that early stopping has not terminated the run.

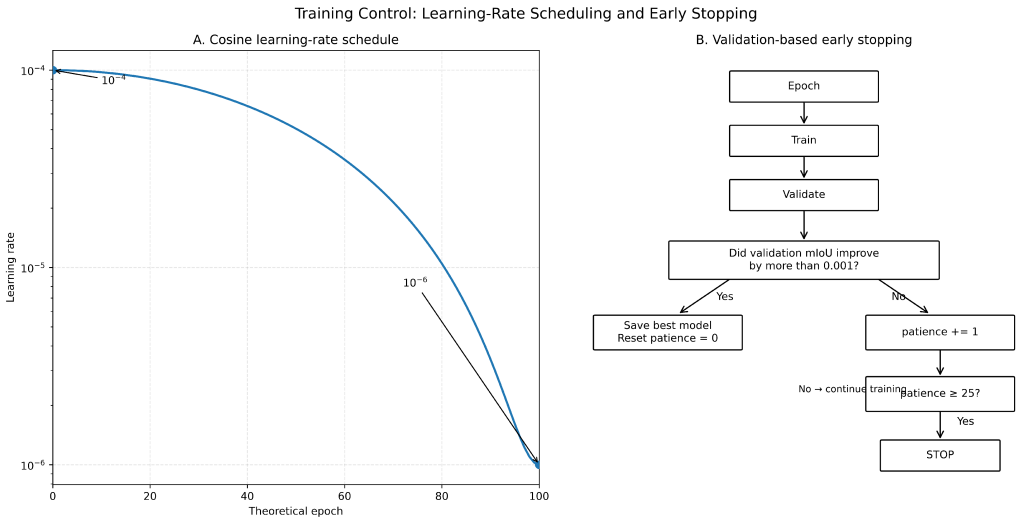


Figure 5.11: Training control mechanisms used during the deep learning experiments. Panel A illustrates the cosine learning-rate schedule, decaying from an initial rate of 10^{-4} to a minimum of 10^{-6} over a theoretical maximum of 100 epochs. Panel B outlines the validation-based early-stopping logic, which monitors the validation mIoU and terminates training if no improvement greater than 0.001 is observed for 25 consecutive epochs to prevent overfitting.

5.5.9 Backpropagation sequence

For each training batch, the implemented optimisation sequence is:

$$\mathbf{X} \xrightarrow{f_\theta} \mathbf{Z} \xrightarrow{\text{mask}} \mathbf{Z}_v \xrightarrow{\mathcal{L}_{\text{WCE}}} \mathcal{L} \xrightarrow{\text{backpropagation}} \nabla_\theta \mathcal{L} \xrightarrow{\text{Adam}} \theta'. \tag{5.69}$$

More explicitly, the model produces the complete ten-class logit tensor. The target is reshaped to a categorical integer representation. Pixels belonging to ignored classes are removed from the vectors used for loss calculation. Gradients are reset, the loss is backpropagated, and Adam updates the model parameters.

The training code also calculates patch-level accuracy, macro F1, and mIoU after the forward pass. These metrics are used for monitoring rather than directly optimised by the loss function.

5.5.10 Summary of the final training configuration

Table 5.8 summarises the principal final CC50 training configuration.

Table 5.8: Summary of the principal training configuration used for the final CC50 experiments.

Configuration item	**Value**
Valid scenes	277
Training scenes	249
Validation scenes	28
Training patches	2241
Validation patches	252
Input channels	28
Output classes	10
Network input size	256×256
Scene size before patching	512×512
Patch stride	128
Model	2-D MONAI SwinUNETR
Feature size	48
Batch size	1
Training augmentation	Horizontal and vertical flips
Flip probability	0.5 independently
Loss	Class-weighted cross entropy
Ignored target class	0 (NoData)
Class-weight method	Inverse square-root frequency
Optimizer	Adam
Initial learning rate	1×10^{-4}
Weight decay	0.01
Scheduler	CosineAnnealingLR
$T_{\max}$	100
Minimum learning rate	1×10^{-6}
Maximum epochs	100
Monitored model-selection metric	Validation mIoU
Early-stopping warm-up	5 epochs
Minimum improvement	0.001
Final CC50 patience	25 epochs
Best checkpoint	`best_model.pt`
Random state / principal seed	42

5.6 Accuracy Assessment Metrics

5.6.1 Need for multiple complementary metrics

Accuracy assessment is fundamental in remote-sensing classification because a single summary statistic can hide substantial variation among land-cover classes [39, 40]. This consideration is especially important for the Philippine dataset because the class distribution is highly imbalanced.

If one widespread class occupies a large fraction of the validation pixels, a model can obtain a high overall number of correct pixels while performing poorly on smaller classes. For this reason, the evaluation procedure uses both global and class-specific indicators.

The principal metrics are:

$$\text{OA}, \quad \text{Precision}, \quad \text{Recall}, \quad F_1, \quad \text{IoU}, \quad \text{mIoU}, \quad \kappa. \tag{5.70}$$

The confusion matrix is the basis for the class-specific metrics.

5.6.2 Confusion matrix

For K classes, the confusion matrix is

$$\mathbf{M} \in \mathbb{N}^{K \times K}, \tag{5.71}$$

where element

$$M_{ij} \tag{5.72}$$

contains the number of pixels whose reference class is i and predicted class is j.

In the confusion matrices used in this thesis:

$$\text{rows} = \text{reference classes} \tag{5.73}$$

and

$$\text{columns} = \text{predicted classes}. \tag{5.74}$$

The diagonal elements therefore represent correctly classified pixels.

For class c,

$$TP_c = M_{cc}, \tag{5.75}$$

$$FN_c = \sum_j M_{cj} - M_{cc}, \tag{5.76}$$

and

$$FP_c = \sum_i M_{ic} - M_{cc}. \tag{5.77}$$

5.6.3 Overall Accuracy

Overall Accuracy is the fraction of valid evaluation pixels assigned the correct class:

$$\mathrm{OA} = \frac{\sum_c M_{cc}}{\sum_i \sum_j M_{ij}}. \tag{5.78}$$

OA is intuitive and remains useful for summarising total map agreement. However, it is strongly influenced by the most frequent classes. It is therefore interpreted together with macro and class-wise metrics rather than used as the sole performance indicator.

5.6.4 Precision and User's Accuracy

The precision of class c is

$$\mathrm{Precision}_c = \frac{TP_c}{TP_c + FP_c}. \tag{5.79}$$

Precision answers the question:

> Of all pixels predicted as class c, what fraction actually belongs to class c in the reference?

In thematic remote-sensing terminology, this is closely related to User's Accuracy. A low precision indicates substantial commission into the predicted class.

5.6.5 Recall and Producer's Accuracy

Recall is

$$\mathrm{Recall}_c = \frac{TP_c}{TP_c + FN_c}. \tag{5.80}$$

It answers:

Of all reference pixels belonging to class c, what fraction was recovered by the model?

In land-cover accuracy assessment, recall corresponds to Producer's Accuracy under the row-reference convention used here. A low recall indicates that a substantial part of the reference class is omitted from the predicted class.

5.6.6 F1 score

The F1 score combines precision and recall through their harmonic mean,

$$F_{1,c} = 2\frac{\text{Precision}_c \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c}. \tag{5.81}$$

F1 becomes high only when both precision and recall are high. It is therefore useful in the present project because strong performance on only one side of the commission–omission trade-off cannot produce a high F1 score.

The macro F1 is

$$F_{1,\text{macro}} = \frac{1}{|\mathcal{C}_p|} \sum_{c \in \mathcal{C}_p} F_{1,c}, \tag{5.82}$$

where $\mathcal{C}_p$ contains the classes considered present for the corresponding calculation.

Macro averaging gives each included class equal contribution regardless of its number of pixels. It is therefore more sensitive to minority-class failure than OA.

5.6.7 Intersection over Union

Intersection over Union (IoU), also known as the Jaccard score in related contexts, is defined for class c as

$$\text{IoU}_c = \frac{TP_c}{TP_c + FP_c + FN_c}. \tag{5.83}$$

IoU simultaneously penalises omission and commission and is one of the standard metrics for semantic segmentation.

Mean IoU is

$$\text{mIoU} = \frac{1}{|\mathcal{C}_p|} \sum_{c \in \mathcal{C}_p} \text{IoU}_c. \tag{5.84}$$

mIoU is particularly important in this thesis because it is also the metric used for validation-based checkpoint selection. A model is therefore not selected solely

because it correctly predicts many pixels belonging to dominant classes; it must improve average segmentation performance across the relevant classes.

5.6.8 Cohen's Kappa

Cohen's Kappa measures observed agreement relative to agreement expected from the marginal class distributions [41]. It is calculated as

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \tag{5.85}$$

where

$$p_o = \mathrm{OA} \tag{5.86}$$

is the observed agreement and p_e is the agreement expected from the row and column marginals.

For a confusion matrix with N evaluated pixels,

$$p_e = \frac{1}{N^2} \sum_c \left(\sum_j M_{cj} \right) \left(\sum_i M_{ic} \right). \tag{5.87}$$

Kappa is retained as a complementary statistic because it has a long history in thematic map assessment. It is not used alone to judge segmentation quality, and the interpretation in Chapter 6 places greater emphasis on the confusion matrix, class-specific precision and recall, F1, and IoU.

5.6.9 Row-normalized confusion matrix

In addition to raw confusion-matrix counts, a row-normalised matrix is generated as

$$\tilde{M}_{ij} = \frac{M_{ij}}{\sum_k M_{ik}}. \tag{5.88}$$

Every non-empty row therefore sums to one. Under the reference-row convention,

$$\tilde{M}_{ii} \tag{5.89}$$

corresponds to the recall of class i.

The off-diagonal elements make systematic confusion patterns directly visible. For example, the matrix can reveal whether a vegetation class is predominantly confused with another vegetation class or whether wetland pixels are being assigned to Water. These patterns are examined in Chapter 6 rather than interpreted here.

5.6.10 Training-time validation metrics

During training, metrics are calculated on the 256×256 validation patches. For each batch, ignored target pixels are removed and the following quantities are calculated:

$$\text{loss}, \quad \text{accuracy}, \quad F_{1,\text{macro}}, \quad \text{mIoU}. \tag{5.90}$$

The epoch-level values stored in the training log are averages over the used validation batches.

For the training-time mIoU implementation, classes present in the ground-truth pixels of the evaluated patch determine the default class set for that local calculation. The resulting patch-level mIoU values are then averaged across the validation batches.

This is the mIoU used by the early-stopping and best-checkpoint logic.

5.6.11 Full-scene reconstruction for final validation

The detailed final evaluation does not simply count all 252 overlapping validation patches independently. Doing so would cause pixels located in overlapping areas to appear multiple times in the final accuracy calculation.

Instead, every original 512×512 validation scene is loaded without pre-patching. The same nine 256×256 windows are extracted, passed through the network, and converted from logits to class probabilities using softmax.

Let

$$\mathbf{P}^{(r)}(x, y, c) \tag{5.91}$$

denote the probability of class c at full-scene location (x, y) contributed by patch r. If location (x, y) is covered by $R(x, y)$ overlapping patches, the reconstructed probability is

$$\bar{P}(x, y, c) = \frac{\sum_{r=1}^{R(x,y)} P^{(r)}(x, y, c)}{R(x, y)}. \tag{5.92}$$

The final reconstructed label is

$$\hat{y}(x, y) = \arg\max_{c} \bar{P}(x, y, c). \tag{5.93}$$

Thus, the principal CC50 evaluation uses uniform arithmetic averaging of the overlapping class probabilities before the final argmax decision.

This reconstruction has two benefits. First, it returns the prediction to the original 512×512 validation-scene geometry. Second, every underlying validation pixel is counted only once in the final global evaluation.

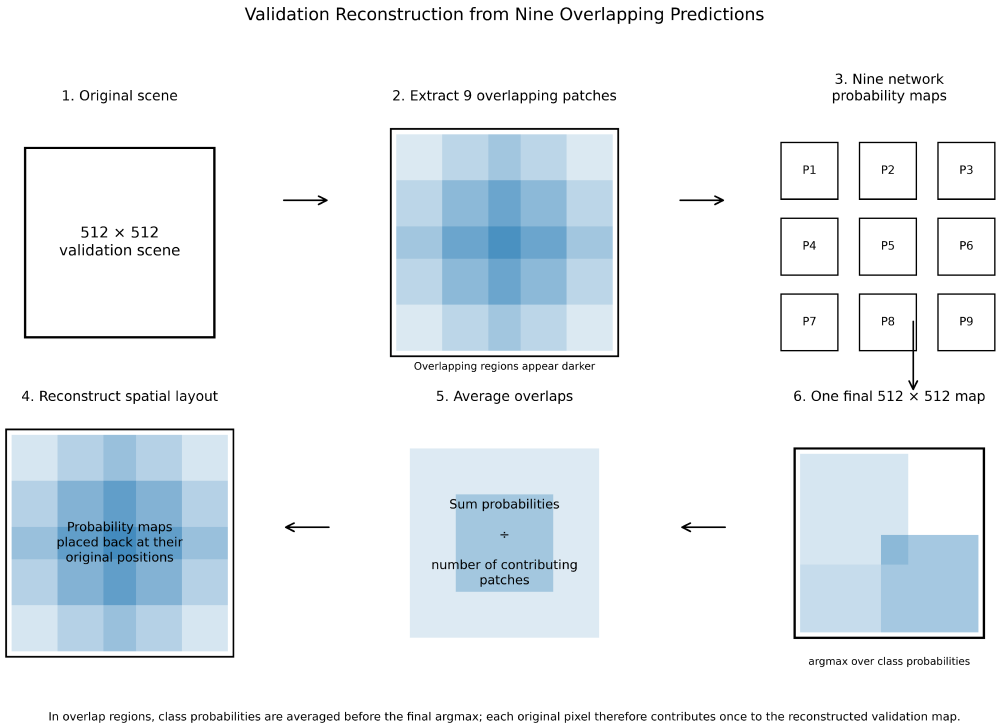


Figure 5.12: Full-scene reconstruction procedure used for final validation. The nine overlapping 256×256 probability maps are remapped to their original spatial coordinates. Overlapping regions are combined via arithmetic averaging of the predicted probabilities before the final argmax decision produces a single 512×512 classification map. This guarantees that every valid reference pixel is counted exactly once during global accuracy assessment.

5.6.12 Global unique-pixel evaluation

After each validation scene is reconstructed, a validity mask is generated from its reference map,

$$V(x,y) = \begin{cases} 1, & Y(x,y) \neq 0, \\ 0, & Y(x,y) = 0. \end{cases} \tag{5.94}$$

Only pixels satisfying $V(x,y) = 1$ contribute to the principal confusion matrix.

The scene-level confusion matrices are then accumulated into a global confusion matrix:

$$\mathbf{M}_{\text{global}} = \sum_{s=1}^{N_{\text{val}}} \mathbf{M}_s. \tag{5.95}$$

The global class-wise precision, recall, F1, IoU, OA, macro metrics, and Cohen's Kappa are computed from this reconstructed unique-pixel evaluation.

This calculation is intentionally distinguished from the training log. The two sets of metrics answer slightly different questions:

$$\begin{aligned} \text{training log:} &\ \text{average performance over validation patches,} \\ \text{final evaluation:} &\ \text{performance over unique pixels of reconstructed scenes.} \end{aligned} \tag{5.96}$$

Therefore, a training-log validation mIoU and a final global reconstructed-scene mIoU are not expected to be numerically identical even when they are obtained from the same checkpoint.

5.6.13 Why OA alone is insufficient

Consider a hypothetical validation scene in which 95% of valid pixels belong to Forest. A classifier predicting Forest almost everywhere could obtain a high OA while completely missing several minority classes. Such a prediction would be unsatisfactory for a multi-class land-cover product despite the apparently high total number of correct pixels.

This is precisely why the assessment reports a set of complementary metrics. OA describes total valid-pixel agreement. Precision describes commission into a class. Recall describes omission from a class. F1 requires balance between precision and recall. IoU evaluates spatial overlap while penalising both types of error. Macro averaging prevents the largest classes from completely dominating the summary score.

The same logic motivates the use of class-wise confusion analysis for the rare classes. Numerical results for Shrubland, Mangrove Forest, Wetland, Grassland,

Built-up, and the more frequent classes are therefore examined separately in Chapter 6.

5.6.14 Metric summary

Table 5.9 summarises the interpretation of the principal metrics.

Table 5.9: Principal metrics used for evaluation of the Philippine land-cover segmentation model.

Metric	Expression	Interpretation
OA	$\frac{\sum_c TP_c}{N}$	Fraction of all valid pixels classified correctly
Precision	$\frac{TP}{TP+FP}$	Reliability of pixels assigned to a class; related to User's Accuracy
Recall	$\frac{TP}{TP+FN}$	Fraction of the reference class recovered; Producer's Accuracy
F1	$\frac{2PR}{P+R}$	Harmonic balance of precision and recall
IoU	$\frac{TP}{TP+FP+FN}$	Overlap between predicted and reference class while penalising commission and omission
mIoU	Mean class IoU	Class-balanced summary of segmentation overlap
Cohen's κ	$\frac{p_o-p_e}{1-p_e}$	Agreement relative to that expected from marginal distributions

5.7 Chapter Summary

This chapter described the complete deep learning methodology used to transform the Sentinel-1 feature dataset into dense Philippine land-cover predictions.

The classification task is formulated as supervised semantic segmentation. Every valid sample combines a MOLCA-derived reference raster with 28 Sentinel-1 VH feature rasters representing seven feature types across four temporal groups. The scene tensors are resized to 512×512 pixels and divided into nine overlapping 256×256 network samples using a stride of 128 pixels.

The model is implemented as a two-dimensional MONAI SwinUNETR with 28 input channels, ten output classes, and a feature size of 48. This implementation belongs to the broader U-shaped Swin-transformer segmentation family on which

the methodology of Russo et al. [9] is based. The architectural discussion therefore distinguishes between the Swin-Unet methodological lineage and the exact SwinUNETR software implementation used in the Philippine experiments.

The 277 valid scenes are separated before patch extraction. Their class-composition vectors are grouped using four K-means clusters, and a stratified scene-level split produces 249 training scenes and 28 validation scenes. After overlapping patch extraction, this corresponds to 2241 training patches and 252 validation patches. Because the split is defined at scene level, overlapping patches from the same parent scene cannot cross the train–validation boundary.

Training uses class-weighted cross-entropy with NoData excluded from valid targets. The class weights are computed from training scenes only using inverse-square-root class frequency. Adam optimisation is used with an initial learning rate of 10^{-4} and weight decay of 0.01, together with cosine learning-rate annealing. The final controlled CC50 configuration allows a maximum of 100 epochs and selects the best checkpoint according to validation mIoU, with early stopping after a five-epoch warm-up and 25 consecutive non-improving epochs.

Two controlled CC50 training protocols are defined. The clean baseline uses the ordinary shuffled training loader while retaining class-weighted loss. The Shrubland sampling experiment additionally assigns a $10\times$ selection weight to every training patch containing at least one Shrubland pixel. Sampling is performed with replacement while preserving the same number of draws per epoch, reference data, validation scenes, model, loss, optimiser, and evaluation procedure.

Finally, the accuracy-assessment framework combines Overall Accuracy, precision, recall, F1, IoU, mIoU, Cohen's Kappa, and confusion-matrix analysis. For final evaluation, overlapping patch probabilities are reconstructed into the original 512×512 validation scenes before metric calculation so that each valid validation pixel contributes only once.

The numerical results generated by this methodology, including model convergence, the confusion matrices, class-specific performance, comparison of the clean and weighted-sampling experiments, qualitative segmentation maps, and analysis of the remaining classification errors, are presented in Chapter 6.

Chapter 6

Results and Discussion

This chapter presents and interprets the main results of the complete Philippine land-cover mapping workflow developed in this thesis. The results are organised according to the sequence of the preceding methodological chapters: construction and revision of the MOLCA-derived reference dataset, selection of the patch dataset, preparation and verification of the Sentinel-1 feature dataset, deep learning model training, qualitative segmentation behaviour, quantitative accuracy assessment, and detailed analysis of the remaining errors.

The chapter also documents an important methodological development that occurred during the project. The first fully trained SwinUNETR model revealed a complete failure to recover Shrubland. Rather than treating this result only as a model-training problem, we returned to the reference data and examined where Shrubland was lost during the MOLCA construction process. That audit showed that Shrubland was nearly absent from the original final reference, which motivated a conservative reference revision based on agreement between NAMRIA and FROM-GLC over previously unclassified pixels.

A connected-component analysis was subsequently used to control the spatial fragmentation of the rescued Shrubland labels. The CC50 version was retained as the main revised reference because it provided an intermediate compromise between the noisier CC10 alternative and the substantially more conservative CC100 version. A new clean baseline was then trained on CC50. Finally, a controlled 10× WeightedRandomSampler experiment was performed to test whether substantially increasing the exposure of Shrubland-containing training patches could resolve the remaining failure.

The results show that the final clean CC50 model provides strong internal validation performance for the dominant land-cover classes, particularly Water, Forest, and Cropland, with an Overall Accuracy of 0.8333 and a global full-scene mean IoU of 0.4339. However, the Shrubland class remains problematic. The 10× sampler changes the model from predicting no Shrubland at all to generating

26,056 Shrubland predictions, but only 39 of these predictions are correct. The sampler therefore increases class-5 prediction frequency without learning a reliable Shrubland decision boundary, while the principal global metrics decrease.

A subsequent diagnostic analysis indicates that the problem cannot be explained by class imbalance alone. The CC50 training set already contains substantial rescued Shrubland supervision, yet the available Sentinel-1 feature representation shows considerable geographic variability in the relationship between Shrubland, Forest, and Grassland. These findings are therefore interpreted as evidence of a combined rare-class, class-separability, and geographic-generalisation problem rather than as the effect of one isolated factor.

6.1 MOLCA Philippines Reference Dataset

6.1.1 Final reference-map coverage before the Shrubland revision

The first important result concerns the composition of the MOLCA Philippines reference generated using the agreement procedure described in Chapter 3. Across the 60 operational Sentinel-2 tiles, the reference contained more than 5.79 billion valid classified pixels, while approximately 1.74 billion pixels remained unclassified/AnA.

Table 6.1 summarizes the class distribution before the later Shrubland-specific revision.

Table 6.1: Class distribution of the original final MOLCA Philippines reference across the 60 operational Sentinel-2 tiles before the conservative Shrubland revision.

ID	Class	Number of pixels
1	Built-up	71,247,316
2	Cropland	505,169,006
3	Forest	1,778,117,338
4	Grassland	117,238,811
5	Shrubland	16,699
6	Wetland	21,731,262
7	Water	3,279,717,034
8	Bareland	4,226,300
9	Mangrove Forest	20,388,898
–	Valid labelled pixels	5,797,852,664
–	Unclassified / AnA pixels	1,739,537,892
–	Total pixels	7,537,390,556

The distribution is strongly imbalanced. Water is the most abundant class, followed by Forest and Cropland. In contrast, Shrubland occupies only 16,699 pixels across the complete 60-tile reference. This corresponds to approximately.

$$\frac{16\,699}{5\,797\,852\,664} \times 100 \approx 0.00029\% \tag{6.1}$$

of the valid labelled pixels.

Shrubland was also geographically restricted: it occurred in only six of the 60 operational tiles. The rarity was therefore present at two levels: Shrubland contributed very few pixels and occurred in very few geographic areas.

This result became particularly important after the first full deep learning experiment. The model's complete failure to recover Shrubland could not be interpreted adequately without considering the reference distribution from which the training labels had been obtained.

6.1.2 Evidence from the source-to-final Shrubland audit

The rare-class audit performed after the first full model training examined whether Shrubland was genuinely absent from the available source products or whether it was being removed by the agreement hierarchy used to construct the reference map.

The audit showed that Shrubland existed in the source products, but very little of it survived as final MOLCA Shrubland. In particular, pixels mapped as

Shrubland by NAMRIA were overwhelmingly assigned to other final categories, especially Forest, or remained unclassified.

An even stronger indication was obtained by considering locations where both NAMRIA and FROM-GLC indicated Shrubland. These locations represent a conservative agreement condition involving two independent source products. Nevertheless, the original final MOLCA still retained only a very small fraction of these pixels as Shrubland.

The observed transition pattern is consistent with the hierarchical MOLCA logic described in Chapter 3. A Shrubland pixel that fails the initial FROM-GLC–GlobeLand30 base agreement can remain unclassified during the early stages. If FNF subsequently supports Forest, the Forest evidence stage may classify that previously unresolved pixel as Forest. Since the original workflow did not contain a corresponding Shrubland-specific rescue stage, Shrubland could be progressively depleted without any explicit rule stating that Shrubland should be replaced by Forest.

This distinction is important. The audit does not show that the Forest rule was "incorrect." It shows that the combination of a strict base agreement, class-specific thematic rescue stages, and the lack of an equivalent Shrubland evidence stage produced an asymmetric effect on the final class distribution.

6.1.3 Conservative Shrubland rescue

Following the reference audit and discussions with the supervisors, a conservative Shrubland rescue rule was evaluated:

$$\begin{aligned} &\text{Original MOLCA}(p) = \text{unclassified} \\ &\wedge\ \text{NAMRIA}(p) = \text{Shrubland} \\ &\wedge\ \text{FROM-GLC}(p) = \text{Shrubland} \\ &\Longrightarrow \text{MOLCA}_{v2}(p) = \text{Shrubland}. \end{aligned} \tag{6.2}$$

No pixel that had already received a valid land-cover class was overwritten. The unfiltered rule recovered.

$$5\,588\,058 \tag{6.3}$$

Shrubland pixels. The total Shrubland population consequently increased from

$$16\,699 \tag{6.4}$$

to

$$5\,604\,757 \tag{6.5}$$

pixels.

All other valid land-cover classes remained unchanged. The increase in Shrubland corresponded exactly to a decrease in previously unclassified pixels.

This result confirmed that the rescue operation behaved conservatively with respect to the existing MOLCA classes. However, a second problem became apparent when the spatial structure of the newly rescued pixels was analysed.

6.1.4 Fragmentation of the unfiltered rescue

The 5,588,058 rescued pixels were distributed across approximately.

$$1\,654\,059 \tag{6.6}$$

connected components.

This very high number of components indicated strong spatial fragmentation. In several important tiles, the median connected-component size was only one pixel. Consequently, using the complete unfiltered rescue directly for training would have introduced a large number of isolated or very small class-5 objects.

The result illustrates why total recovered pixel count alone was not a sufficient criterion for accepting the revised reference. A reference map can contain many additional class pixels while still being spatially unsuitable for training if those pixels consist mainly of isolated speckle-like components.

Connected-component filtering was therefore evaluated.

6.1.5 CC10, CC50, and CC100 component filtering

Three principal connected-component thresholds were compared:

$$\mathrm{CC10} \geq 10 \text{ pixels}, \tag{6.7}$$

$$\mathrm{CC50} \geq 50 \text{ pixels}, \tag{6.8}$$

and

$$\mathrm{CC100} \geq 100 \text{ pixels}. \tag{6.9}$$

Table 6.2 summarizes the resulting Shrubland retention.

Table 6.2: Effect of connected-component filtering on the conservative Shrubland rescue.

Version	Rescued pixels retained	Retention of raw rescue	Total Shrubland	Shrubland % of all pixels
Unfiltered	5,588,058	100.00%	5,604,757	0.07436%
CC10	2,404,877	43.04%	2,421,576	0.032128%
CC50	783,561	14.02%	800,260	0.010617%
CC100	398,822	7.14%	415,521	0.005513%

CC10 retained the largest amount of newly rescued Shrubland but remained visually fragmented in several areas. CC100 produced substantially cleaner and more conservative spatial objects but retained only approximately 7.14% of the unfiltered rescue.

CC50 retained approximately 14.02% of the raw rescued pixels. It therefore reduced the strongest pixel-level fragmentation while retaining almost twice as much rescued Shrubland as CC100.

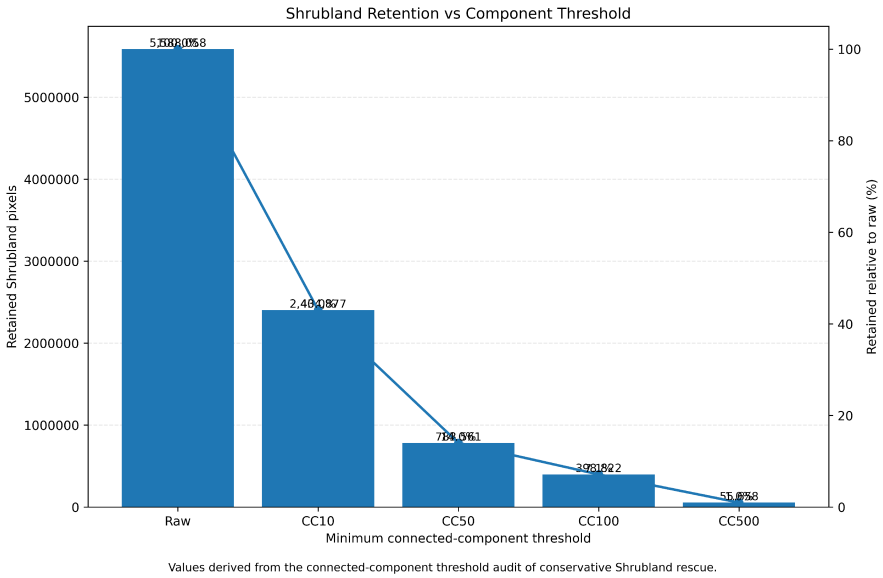


Figure 6.1: Quantitative effect of connected-component filtering on the conservative Shrubland rescue. The dual-axis plot illustrates the absolute number of retained pixels (bars) and the retention percentage relative to the raw unfiltered rescue (line) across progressively stricter thresholds. This confirms the steep initial loss of fragmented pixels and supports the selection of CC50 as a balanced intermediate threshold.

The visual comparison of the original, CC10, CC50, and CC100 maps supported this intermediate interpretation.

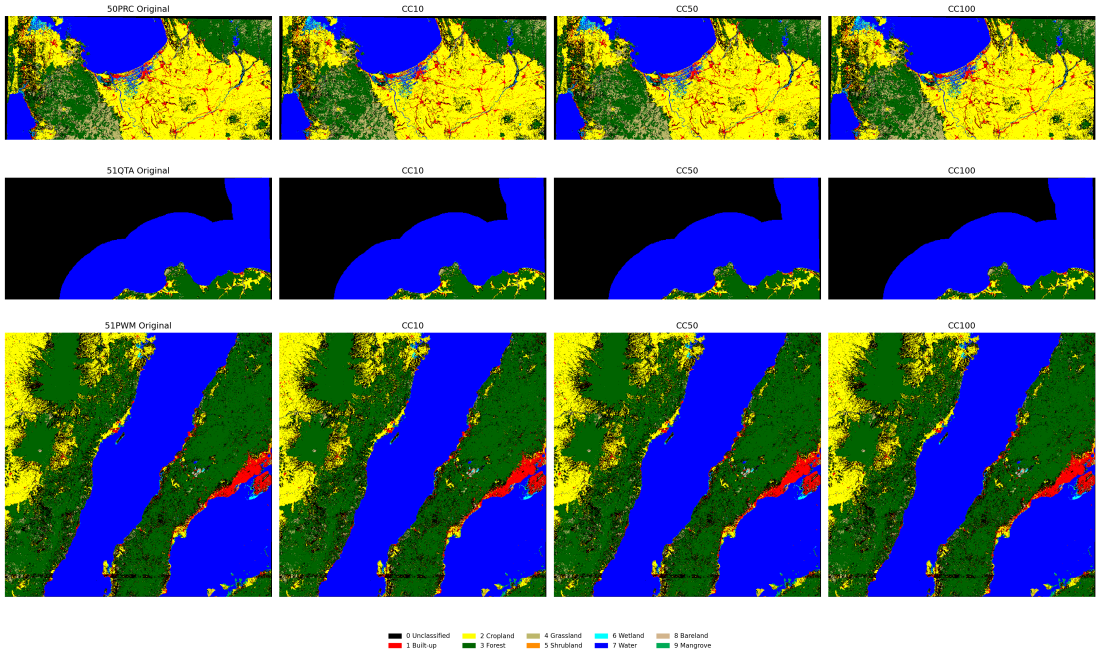


Figure 6.2: Visual comparison of the original MOLCA reference and the CC10, CC50, and CC100 connected-component filtering variants for three representative scenes (50PRC, 51QTA, and 51PWM). The comparison illustrates the progressive removal of small isolated Shrubland components as the minimum connected-component threshold increases.

6.1.6 Selection of CC50 as the principal revised reference

Based on the quantitative component-retention analysis and visual inspection, CC50 was retained as the principal revised reference. This decision was subsequently reviewed with the project supervisors. Luigi Russo agreed that the unfiltered version was too fragmented, CC10 remained relatively noisy, and CC100, although cleaner, was probably too conservative. CC50 was therefore adopted as the main revised reference, while CC100 was preserved as a stricter comparison product.

The CC50 product contains

$$800\,260 \tag{6.10}$$

Shrubland pixels over the 60-tile reference, of which

$$783\,561 \tag{6.11}$$

were introduced by the filtered conservative rescue and

$$16\,699 \tag{6.12}$$

were inherited from the original MOLCA product.

Importantly, the choice of CC50 was made before retraining the final model. The threshold was therefore not selected retrospectively to maximise model accuracy. It was selected from reference-map evidence based on label conservativeness, spatial coherence, and retained class coverage.

6.1.7 Implications of the reference-map result

The reference-map analysis produces three important conclusions.

First, agreement-based reference generation can produce class-specific imbalances that are much stronger than the class distribution in any individual source product. This is particularly important when some classes benefit from dedicated thematic rescue rules while others do not.

Second, simply recovering all available source-agreement pixels is not necessarily desirable. The unfiltered rescue recovered millions of pixels, but the connected-component analysis showed that many were extremely fragmented.

Third, the CC50 revision substantially increased available Shrubland supervision without overwriting existing valid MOLCA classes. This made it possible to separate two questions experimentally:

1. Was the original model failure mainly caused by almost absent Shrubland labels?

2. Would the model learn Shrubland once a larger conservative set of class-5 labels was introduced?

The deep learning experiments reported in Sections 6.4 and 6.6 address these questions.

6.2 Selected Patch Dataset

6.2.1 Final geographical sample

The patch-selection procedure of Chapter 4 produced a final set of 280 geographical patches distributed across 60 Sentinel-2 tiles.

The selected sample was intentionally not a simple random sample of the national reference. It was designed to retain broad geographic coverage while also increasing the representation of rare and informative land-cover conditions.

The final selection therefore served the requirements of supervised deep learning rather than statistical estimation of national land-cover area.

The multi-scale patch visualisations presented in Chapter 4 show that the selected areas cover a wide range of Philippine landscape configurations, including coastal areas, agricultural mosaics, forested interiors, water-dominated environments, wetlands, built-up regions, and rare-class locations.

6.2.2 Transition from 280 selected patches to 277 valid DL scenes

The selected geographical dataset initially contained 280 patches. During the Sentinel-1 input verification, the three selected patches belonging to tile `51QUC` were found to lack valid 2019 VH Sentinel-1 data required by the final model input.

These samples were therefore excluded before deep learning training.

The final number of valid scene associations was

$$280 - 3 = 277. \tag{6.13}$$

This resulted in:

$$249 \text{ training scenes} \tag{6.14}$$

and

$$28 \text{ validation scenes.} \tag{6.15}$$

The exclusion was performed before model training rather than attempting to replace the missing SAR values with synthetic information.

6.2.3 CC50 Shrubland representation within the selected scenes

The revised CC50 reference materially changed the Shrubland content of the selected scenes.

In model space, the final training set contains:

$$62\,501 \tag{6.16}$$

Shrubland pixels distributed across 60 training scenes.

The validation set contains:

$$3\,716 \tag{6.17}$$

Shrubland pixels distributed across five validation scenes.

At the 256×256 training-patch level,

$$296 \tag{6.18}$$

of the 2241 training patches contain at least one Shrubland pixel.

The validation set contains 19 Shrubland-containing patches among the 252 overlapping validation patches.

Table 6.3 summarizes the CC50 Shrubland exposure.

Table 6.3: Shrubland representation in the final CC50 deep learning dataset.

Quantity	**Value**
Training scenes	249
Training scenes containing Shrubland	60
Validation scenes	28
Validation scenes containing Shrubland	5
Training patches	2241
Shrubland-containing training patches	296
Validation patches	252
Shrubland-containing validation patches	19
Training Shrubland pixels in model space	62,501
Validation Shrubland pixels in model space	3,716

This distribution is important for the interpretation of the later model failure. Unlike the original MOLCA reference, the revised training data no longer contain only a negligible number of Shrubland labels. Nevertheless, Shrubland remains uncommon relative to the dominant classes and unevenly distributed across geographic scenes.

6.2.4 Reference provenance of the training Shrubland

The final diagnostic analysis traced every model-space training Shrubland pixel to either the original MOLCA reference or the CC50 rescue.

Of the 62,501 training Shrubland pixels:

$$9\,276 \tag{6.19}$$

were inherited from the original MOLCA, corresponding to

$$14.84\%, \tag{6.20}$$

while

$$53\,225 \tag{6.21}$$

were introduced by the CC50 rescue, corresponding to

$$85.16\%. \tag{6.22}$$

At the scene level, 12 training scenes contain original-MOLCA Shrubland and 50 contain rescued Shrubland, with two scenes containing both provenance types.

At the overlapping-patch level, 62 patches contain original Shrubland, 237 contain rescued Shrubland, and three contain both.

Table 6.4: Provenance of Shrubland supervision in the CC50 training set.

Training quantity	**Original MOLCA**	**CC50 rescue**
Shrubland pixels	9,276	53,225
Fraction of all training Shrubland	14.84%	85.16%
Scenes containing provenance type	12	50
Overlapping patches containing provenance type	62	237

This result becomes central in Section 6.6.7. The failure of the model on the rescued validation scenes cannot be explained simply by claiming that rescued Shrubland was absent from training. The majority of the training Shrubland supervision is already rescue-derived.

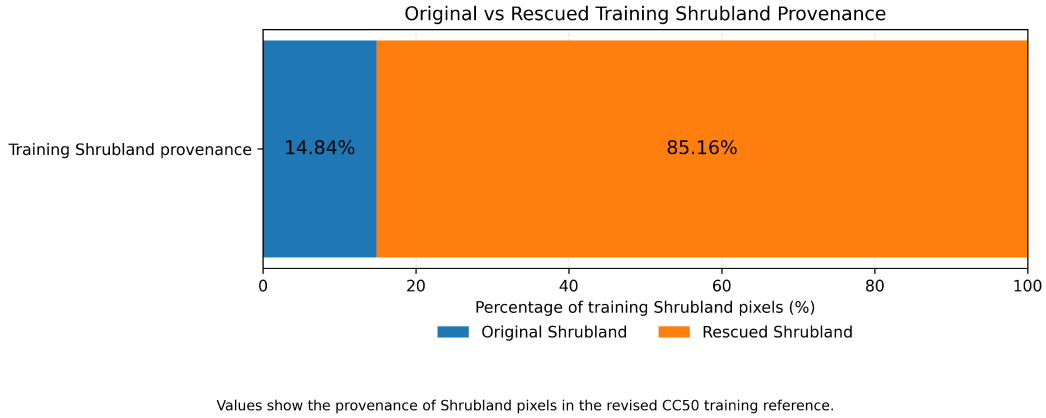


Figure 6.3: Provenance of Shrubland pixels in the revised CC50 training reference. Only 14.84% of training Shrubland pixels originated from the original reference, whereas 85.16% were introduced by the conservative rescue procedure. This illustrates the strong dependence of the final Shrubland training set on the revised reference generation strategy.

6.3 Sentinel-1 Feature Dataset

6.3.1 Completeness of the final Sentinel-1 feature set

After removal of the three invalid 51QUC samples, every one of the 277 valid scenes contained the complete 28-channel Sentinel-1 feature set used by the final model.

The final number of usable Sentinel-1 feature files was therefore

$$277 \times 28 = 7\,756. \tag{6.23}$$

The indexing verification reported one ground-reference raster and 28 associated Sentinel-1 feature rasters for every retained scene, with no remaining skipped samples.

The channel sequence was also audited across all 277 scenes and found to be consistent:

$$\text{SI, LEE, MAX, MIN, MAXMIN, MEAN, MEDIAN} \tag{6.24}$$

repeated for each of the four temporal groups.

This verification is important because an ordering error would alter the physical meaning associated with each input channel while leaving the tensor shape unchanged.

6.3.2 Spatial alignment and dimensions

The QGIS and local post-processing checks presented in Chapter 4 confirmed that Sentinel-1 features and their corresponding ground-reference rasters were brought to a common 549×549 local product before the deep learning loader resized them to 512×512.

For the explicit verification example `51PTR_R09_C15`, both the ground-reference raster and corresponding Sentinel-1 SI raster had dimensions of 549×549 pixels and effectively coincident projected extents.

This supports the spatial consistency of the input/label pairs entering the deep learning pipeline.

6.3.3 Information content of the seasonal feature representation

The four temporal groups provide visibly different VH backscatter representations of the same geographic areas. The seven derived feature types also emphasise different aspects of spatial variation.

This is consistent with the broader motivation of the spatio-temporal Sentinel-1 methodology developed by Russo et al. [9], in which seasonal synthesis is used to represent temporal behaviour without requiring a dense acquisition sequence as separate model inputs.

The strong performance subsequently obtained for Water, Forest, and Cropland indicates that the complete 28-channel representation contains substantial class-discriminating information. However, class performance is not uniform, and the Shrubland diagnostics show that strong performance for major classes does not guarantee stable recovery of every rare vegetation category across all validation scenes.

6.4 Model Training Results

6.4.1 End-to-end one-epoch verification

Before the first full training run, a complete one-epoch experiment was used to verify that the adapted Philippine pipeline worked end-to-end.

The local verification confirmed:

1. correct indexing of the 277 valid scenes;
2. correct association of 28 Sentinel-1 inputs with each reference;
3. correct ten-class model output;

4. successful class-weighted loss calculation;

5. successful backward propagation;

6. validation execution;

7. successful checkpoint generation.

The one-epoch run was therefore a software and data-pipeline verification, not a final model-performance experiment.

On the Linux GPU workstation, the one-epoch verification produced approximately:

$$\text{Train accuracy} \approx 65.6\%, \tag{6.25}$$

$$\text{Validation accuracy} \approx 73.2\%, \tag{6.26}$$

and validation mIoU of approximately 0.308.

These preliminary values showed that the model could learn a meaningful signal, but no scientific conclusion concerning final land-cover quality was drawn from the one-epoch run.

6.4.2 First complete pre-CC50 baseline

The first complete Philippine training run used the original reference before the later CC50 revision.

The maximum configured duration was 100 epochs. With a patience of 15 epochs, early stopping terminated training at epoch 58. The best checkpoint was obtained at epoch 43 with a patch-level validation mIoU of approximately

$$0.3742. \tag{6.27}$$

The full baseline produced:

$$\text{OA} \approx 0.829, \tag{6.28}$$

$$\kappa \approx 0.764, \tag{6.29}$$

and

$$F_{1,\text{macro}} \approx 0.538. \tag{6.30}$$

This model demonstrated that the adapted SwinUNETR pipeline could produce a useful Philippine land-cover segmentation. Water, Forest, and Cropland were already strong classes. However, Shrubland recall was zero.

The complete Shrubland collapse was the key result of this first baseline. It triggered the reference audit reported in Section 6.1 and changed the next stage of the project from model tuning to reference investigation.

6.4.3 Chronology of the principal experiments

The principal experimental stages are summarised in Table 6.5.

Table 6.5: Chronology of the principal deep learning experiments.

Experiment	**Reference**	**Purpose**
One-epoch verification	Original MOLCA	Confirm complete pipeline compatibility, loss, validation, and checkpointing.
First full baseline	Original MOLCA	Establish first trained Philippine model and identify class-specific weaknesses.
Reference audit	Original MOLCA + source products	Determine why Shrubland was almost absent from final reference.
CC10/CC50/ CC100 comparison	Revised Shrubland rescue	Identify a spatially conservative revised reference.
CC50 clean baseline	CC50	Measure effect of revised labels without Shrubland-specific sampling.
CC50 sampler 10×	CC50	Test whether greatly increased exposure to Shrubland-containing patches improves class-5 learning.
Post-training diagnostic	CC50	Compare the five Shrubland validation scenes and investigate reference provenance, posterior probabilities, and cross-scene prediction behaviour.

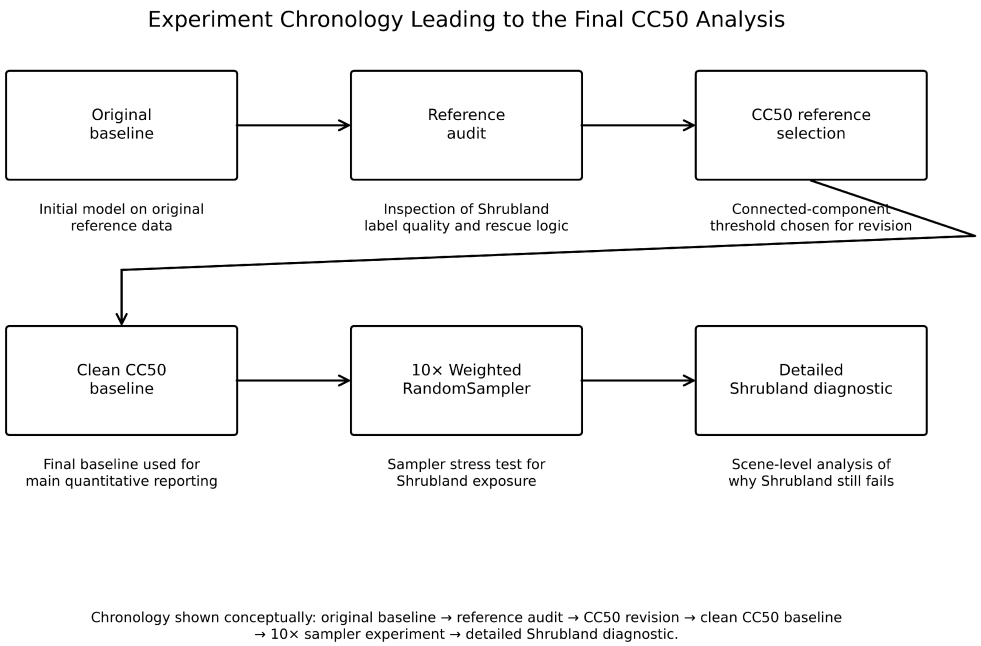


Figure 6.4: Chronology of the main experiments and diagnostic steps leading to the final CC50 analysis. The workflow began with the original baseline, followed by a reference audit and the selection of the CC50 connected-component threshold. A clean CC50 baseline was then established, after which a 10× WeightedRandomSampler experiment was tested. Finally, a detailed Shrubland diagnostic was conducted to investigate the remaining failure cases.

6.4.4 CC50 clean-baseline convergence

The revised CC50 clean baseline used the same overall model design described in Chapter 5, without a Shrubland-specific sampler.

Training was allowed to continue for up to 100 epochs with the final 25-epoch patience configuration.

Early stopping occurred at:

$$E_{\text{stop}} = 78. \tag{6.31}$$

The best checkpoint occurred at:

$$E_{\text{best}} = 53. \tag{6.32}$$

The best patch-level validation mIoU was

$$0.3696774. \tag{6.33}$$

The later full-scene reconstructed evaluation produced a higher numerical mIoU of 0.433949 because this is a different aggregation procedure: the training monitor averages patch-level validation measurements, whereas the final evaluator reconstructs complete validation scenes and counts every valid scene pixel once.

The two values should therefore not be interpreted as inconsistent estimates of exactly the same aggregation.

6.4.5 CC50 clean-baseline global results

The final CC50 clean baseline, using the complete 28-channel Sentinel-1 input configuration described in Chapter 5, was evaluated over

$$6\,174\,226 \tag{6.34}$$

valid reference pixels.
Of these,

$$5\,144\,963 \tag{6.35}$$

were classified correctly.
The resulting overall metrics were:

Table 6.6: Global full-scene unique-pixel validation results for the CC50 clean baseline.

Metric	**Value**
Overall Accuracy	0.833297
Macro precision	0.549740
Macro recall	0.583169
Macro F1	0.546032
Mean IoU	0.433949
Cohen's κ	0.770260

The results establish the CC50 clean model as a strong internal baseline for the Philippine dataset. Its OA is approximately 83.3%, and the Kappa value indicates substantial agreement relative to the marginal class distributions. More importantly for the imbalanced problem, macro F1 remains above 0.54 despite the complete failure of Shrubland.

The result also confirms the importance of reporting more than OA. Class-specific results vary substantially, as shown in Section 6.6.

6.4.6 CC50 10× sampler training result

The controlled sampler experiment increased the expected fraction of Shrubland-containing training draws from approximately

$$13.21\% \tag{6.36}$$

to

$$60.35\%. \tag{6.37}$$

The sampler experiment stopped earlier than the clean baseline:

$$E_{\text{stop}} = 45, \tag{6.38}$$

with the best checkpoint at

$$E_{\text{best}} = 20. \tag{6.39}$$

Its best patch-level validation mIoU was

$$0.3524742, \tag{6.40}$$

which is lower than the clean-baseline value of 0.3696774.
The full-scene global results are shown in Table 6.7.

6.4.7 Direct clean-versus-sampler comparison

Table 6.7: Controlled comparison of the CC50 clean baseline and the 10× Shrubland-patch WeightedRandomSampler experiment.

Metric	**CC50 clean**	**Sampler 10×**	**Sampler − clean**
Best patch-level val. mIoU	0.369677	0.352474	-0.017203
Overall Accuracy	0.833297	0.803817	-0.029480
Macro precision	0.549740	0.505413	-0.044327
Macro recall	0.583169	0.551970	-0.031199
Macro F1	0.546032	0.512222	-0.033810
Full-scene mean IoU	0.433949	0.403811	-0.030138
Cohen's κ	0.770260	0.732253	-0.038007

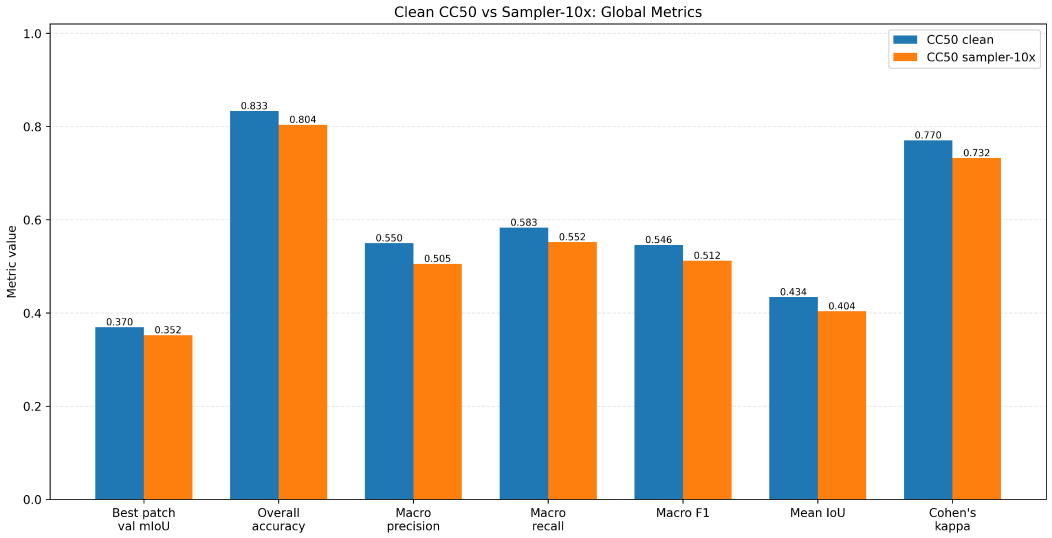


Figure 6.5: Comparison of global performance metrics between the CC50 clean baseline and the 10× Shrubland-patch sampler experiment. The bar chart visually confirms that artificially increasing Shrubland training exposure degrades Overall Accuracy, macro F1, and mean IoU globally, indicating that patch-level exposure alone does not resolve the rare-class learning difficulty.

Every reported global metric decreases in the sampler experiment.

The OA decreases by approximately 2.95 percentage points, while macro F1 decreases by approximately 0.034 and full-scene mIoU by approximately 0.030.

The experiment therefore does not support the interpretation that stronger Shrubland-patch exposure improves the segmentation model as a whole.

The central question is whether this global deterioration is compensated by a meaningful improvement in Shrubland. The class-specific results show that it is not.

6.5 Segmentation Maps

6.5.1 Role of qualitative map inspection

Numerical metrics provide a compact summary of classification quality, but they cannot fully describe the spatial structure of the errors.

Qualitative map inspection was therefore used throughout model development to assess:

1. preservation of coastlines and large water bodies;

2. spatial coherence of Forest and Cropland;
3. recovery of smaller minority-class regions;
4. boundary smoothing;
5. fragmentation;
6. behaviour in heterogeneous coastal scenes;
7. cases in which high OA is produced primarily by a dominant class.

The preserved GT–prediction–error-map examples currently available in the project archive correspond to the first full pre-CC50 baseline. They are included here as evidence of model behaviour during the development cycle. They should not be interpreted as visual outputs from the later CC50 clean checkpoint.

The final CC50 quantitative conclusions in this chapter are based on the epoch-53 reconstructed full-scene evaluation.

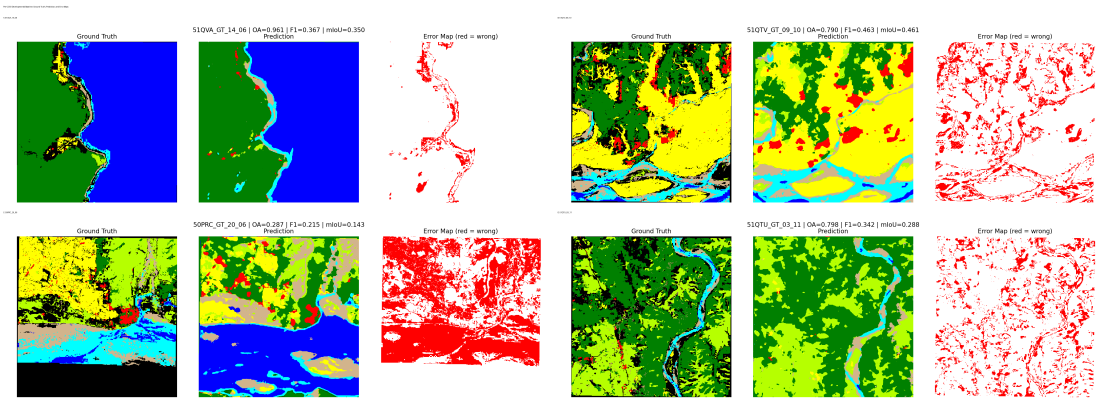


Figure 6.6: Qualitative land-cover predictions from the pre-CC50 developmental baseline for four representative validation scenes. Each example shows the reference ground truth, model prediction, and corresponding error map, where red pixels indicate disagreement between the prediction and reference. These maps are included as developmental qualitative results and should not be interpreted as outputs from the final CC50 clean baseline.

6.5.2 Strong coastal example: 51QVA_14_06

For the first full baseline, the coastal scene `51QVA_14_06` (shown in Panel A of Figure 6.6) provided one of the stronger qualitative results.

Its metrics were approximately:

$$\mathrm{OA} = 0.961, \tag{6.41}$$

$$F_{1,\mathrm{macro}} = 0.367, \tag{6.42}$$

and

$$\mathrm{mIoU} = 0.350. \tag{6.43}$$

The coastline, the large contiguous water body, and the broad terrestrial vegetation structure were reproduced well. As seen in the corresponding error map, most remaining errors were concentrated tightly around narrow class boundaries and highly localised heterogeneous zones.

Interestingly, the final CC50 clean evaluation of the same validation scene remained highly accurate, and its balanced metrics improved substantially:

$$\mathrm{OA} = 0.9683, \tag{6.44}$$

$$F_{1,\mathrm{macro}} = 0.6198, \tag{6.45}$$

and

$$\mathrm{mIoU} = 0.5195. \tag{6.46}$$

Thus, although the exact prediction boundaries shifted as the model learned from the revised dataset, this specific geographic area remained one of the most reliable and well-performing validation examples throughout the study.

6.5.3 Mixed landscape with strong class-level performance

Scene `51QTV_09_10` contains a heterogeneous mixture of Cropland, Forest, Grassland, and Water.

In the first full baseline, it achieved approximately:

$$\mathrm{OA} = 0.790, \tag{6.47}$$

$$F_{1,\mathrm{macro}} = 0.463, \tag{6.48}$$

and

$$\mathrm{mIoU} = 0.461. \tag{6.49}$$

The qualitative result was important because the model reproduced multiple classes rather than simply assigning the dominant category.

For the final CC50 clean model, the reconstructed-scene values were:

$$\mathrm{OA} = 0.8061, \tag{6.50}$$

$$F_{1,\mathrm{macro}} = 0.5596, \tag{6.51}$$

and

$$\mathrm{mIoU} = 0.4353. \tag{6.52}$$

The result confirms that this scene remains a useful example of relatively balanced multi-class segmentation.

6.5.4 Why OA alone can be misleading: 51PWM_09_03

The first-baseline scene `51PWM_09_03` demonstrated the limitation of Overall Accuracy particularly clearly.

The early map obtained:

$$\mathrm{OA} \approx 0.946 \tag{6.53}$$

but only

$$F_{1,\mathrm{macro}} \approx 0.143. \tag{6.54}$$

The scene was dominated by Forest. Correct prediction of the majority class therefore produced very high pixel accuracy despite weak recovery of smaller classes.

In the final CC50 clean evaluation, the same scene obtained:

$$\mathrm{OA} = 0.9410, \tag{6.55}$$

$$F_{1,\mathrm{macro}} = 0.4411, \tag{6.56}$$

and

$$\mathrm{mIoU} = 0.3848. \tag{6.57}$$

The example remains useful as a demonstration that OA should be interpreted alongside class-balanced metrics.

6.5.5 Difficult heterogeneous coastal scene: 50PRC_20_06

The mixed coastal scene `50PRC_20_06` was the hardest qualitative example in the first baseline.

The initial full-baseline result was approximately:

$$\mathrm{OA} = 0.287. \tag{6.58}$$

The scene contains a difficult combination of Cropland, Grassland, Bareland, Water, and a large unclassified region.

The final CC50 clean model improved the scene relative to that earlier developmental result, although it remained one of the weakest validation areas:

$$\mathrm{OA} = 0.4616, \tag{6.59}$$

$$F_{1,\mathrm{macro}} = 0.3859, \tag{6.60}$$

and

$$\mathrm{mIoU} = 0.2625. \tag{6.61}$$

This supports the broader observation that heterogeneous coastal and transition environments remain more difficult than scenes dominated by a large, structurally distinct land-cover class.

6.5.6 Variation across the final CC50 validation scenes

The final CC50 per-scene evaluation shows substantial variability across the 28 validation scenes.

The highest OA values include:

$$0.9744 \tag{6.62}$$

for `51NVJ_09_04`,

$$0.9737 \tag{6.63}$$

for `51PXQ_17_03`,
and

$$0.9683 \tag{6.64}$$

for `51QVA_14_06`.

However, high OA does not necessarily imply the highest class-balanced performance.

The highest per-scene mIoU in the final CC50 clean evaluation is

$$0.5195 \tag{6.65}$$

for `51QVA_14_06`, followed by

$$0.4540 \tag{6.66}$$

for `51QTU_01_10` and

$$0.4353 \tag{6.67}$$

for `51QTV_09_10`.

Table 6.8 provides representative examples.

Table 6.8: Selected scene-level results for the final CC50 clean baseline.

Scene	**OA**	**Macro F1**	**mIoU**
51QVA_14_06	0.9683	0.6198	0.5195
51QTU_01_10	0.8226	0.5995	0.4540
51QTV_09_10	0.8061	0.5596	0.4353
51QVV_12_03	0.9207	0.5125	0.4253
51PWM_09_03	0.9410	0.4411	0.3848
50PRC_11_05	0.7398	0.2564	0.1988
50PRC_20_06	0.4616	0.3859	0.2625

The spread of these results emphasises that classification difficulty depends strongly on scene composition and landscape heterogeneity.

6.6 Accuracy Assessment

6.6.1 Per-class performance of the final CC50 clean baseline

Table 6.9 reports the complete class-wise full-scene evaluation for the final clean CC50 model.

Table 6.9: Per-class unique-pixel validation results for the final CC50 clean baseline. NoData is excluded from the accuracy assessment.

Class	Support	Predicted	Precision	Recall	F1	IoU
Built-up	173,461	187,193	0.5212	0.5624	0.5410	0.3708
Cropland	1,491,586	1,570,323	0.8424	0.8869	0.8641	0.7607
Forest	2,546,505	2,588,476	0.8879	0.9025	0.8951	0.8102
Grassland	431,429	275,000	0.6154	0.3923	0.4791	0.3150
Shrubland	3,716	0	0.0000	0.0000	0.0000	0.0000
Wetland	197,457	198,534	0.4852	0.4878	0.4865	0.3215
Water	1,128,402	1,158,187	0.9316	0.9561	0.9437	0.8934
Bareland	189,707	148,696	0.4960	0.3888	0.4359	0.2787
Mangrove Forest	11,963	47,817	0.1680	0.6717	0.2688	0.1553

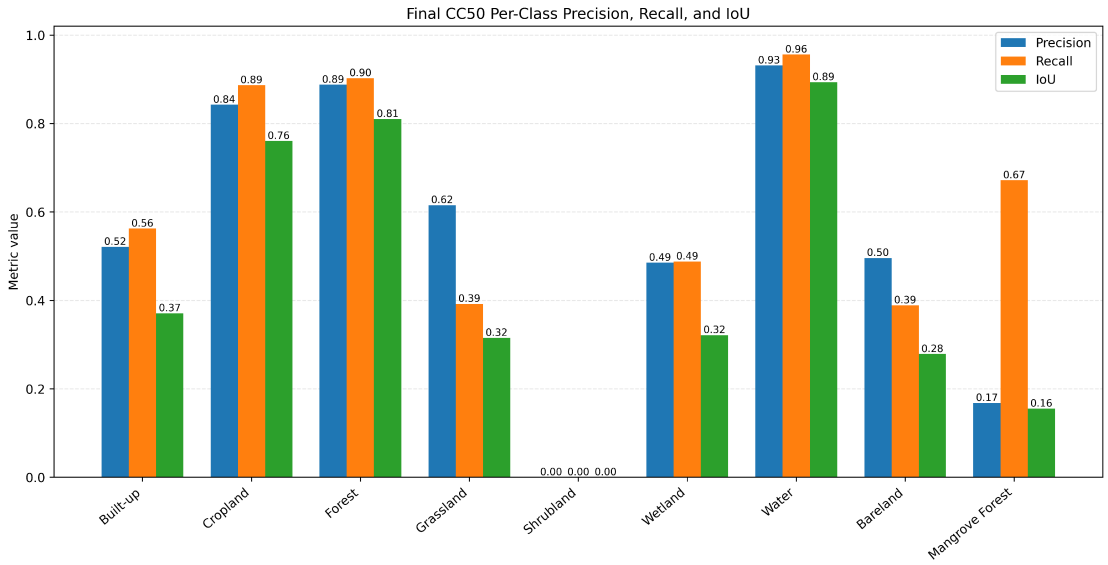


Figure 6.7: Per-class precision, recall, and Intersection over Union (IoU) for the final CC50 clean baseline. The chart visualises the strong performance of dominant classes such as Water, Forest, and Cropland, contrasting with the lower accuracy metrics observed in the minority classes and the complete failure to recover Shrubland.

Three classes are clearly stronger than the rest.
Water reaches:

$$\text{IoU} = 0.8934, \tag{6.68}$$

Forest reaches:

$$\text{IoU} = 0.8102, \tag{6.69}$$

and Cropland reaches:

$$\text{IoU} = 0.7607. \tag{6.70}$$

Their recall values are also high:

$$R_{\text{Water}} = 0.9561, \tag{6.71}$$

$$R_{\text{Forest}} = 0.9025, \tag{6.72}$$

and

$$R_{\text{Cropland}} = 0.8869. \tag{6.73}$$

This indicates that the Sentinel-1 feature representation and SwinUNETR architecture are effective for the large and relatively well-represented classes.

6.6.2 Built-up

Built-up reaches:

$$\text{precision} = 0.5212, \tag{6.74}$$

$$\text{recall} = 0.5624, \tag{6.75}$$

and

$$\text{IoU} = 0.3708. \tag{6.76}$$

The row-normalised confusion matrix shows that approximately 26.0% of true Built-up pixels are classified as Cropland and 15.0% as Forest.

This is consistent with the heterogeneous structure of many Philippine settlements. The reference class includes urban or built surfaces that may occur as compact urban fabric, smaller settlements, road-adjacent development, and scattered structures within agricultural or vegetated landscapes.

At 10 m resolution and using only VH-derived SAR features, these mixed environments can produce substantial within-class variation.

6.6.3 Grassland

Grassland has a precision of 0.6154 but a lower recall of 0.3923.

The difference indicates that when the model predicts Grassland, a substantial fraction of those predictions are correct, but many true Grassland pixels are assigned to other classes.

The largest confusion is Forest:

$$39.10\% \tag{6.77}$$

of true Grassland pixels are predicted as Forest.

A further

$$15.49\% \tag{6.78}$$

are classified as Cropland.

The resulting IoU is 0.3150.

This pattern supports the interpretation that vegetation classes form a more difficult continuum in the available SAR feature space than the dominant Water/Forest/Cropland distinctions.

6.6.4 Wetland

Wetland achieves:

$$\text{precision} = 0.4852, \tag{6.79}$$

$$\text{recall} = 0.4878, \tag{6.80}$$

and

$$\text{IoU} = 0.3215. \tag{6.81}$$

The principal confusion is with Water:

$$30.19\% \tag{6.82}$$

of Wetland reference pixels are predicted as Water.

Another

$$14.74\% \tag{6.83}$$

are predicted as Bareland.

Wetland mapping is inherently sensitive to hydrological state and temporal conditions. Research using combined Sentinel-1 and Sentinel-2 time series has shown the value of complementary spatial, temporal, and spectral information for separating complex wetland classes [42]. The present SAR-only result therefore reflects a difficult classification problem rather than a simple absence of model capacity.

6.6.5 Bareland

Bareland produces:

$$\text{precision} = 0.4960, \tag{6.84}$$

$$\text{recall} = 0.3888, \tag{6.85}$$

and

$$\text{IoU} = 0.2787. \tag{6.86}$$

Its largest confusion in the clean model is Wetland:

$$23.80\%. \tag{6.87}$$

Further confusion occurs with Cropland, Forest, Water, and Grassland.

The relatively broad spread across classes suggests that Bareland is not characterised by one uniquely stable VH pattern in the complete geographic dataset.

6.6.6 Mangrove Forest

Mangrove Forest presents an unusual precision–recall pattern.

Its recall is relatively high:

$$0.6717, \tag{6.88}$$

but its precision is only:

$$0.1680. \tag{6.89}$$

The model therefore recovers approximately two-thirds of the reference Mangrove pixels, but predicts Mangrove much more widely than supported by the reference.

The model predicts 47,817 Mangrove pixels while the validation reference contains only 11,963.

The resulting IoU is

$$0.1553. \tag{6.90}$$

Approximately 24.87% of true Mangrove reference pixels are assigned to Forest. The reverse problem is also important: many non-Mangrove pixels are predicted as Mangrove, explaining the low precision.

Mangrove discrimination is a challenging coastal vegetation problem. Published mangrove-mapping studies frequently combine SAR with optical features and report that the complementary spectral information can improve mangrove mapping compared with SAR alone [43]. The result in this thesis is therefore consistent with a class whose spatial/ecological identity is more specific than can always be inferred from VH backscatter alone.

6.6.7 Shrubland collapse in the CC50 clean model

Despite the substantially revised reference, the clean CC50 model predicts no Shrubland pixels.

The validation reference contains

$$3716 \tag{6.91}$$

Shrubland pixels, but:

$$TP = 0, \tag{6.92}$$

$$FN = 3716, \tag{6.93}$$

and

$$\text{Predicted Shrubland} = 0. \tag{6.94}$$

The corresponding precision, recall, F1, and IoU are all zero.
However, the confusion is not randomly distributed.
Of the true Shrubland pixels:

$$2714 \tag{6.95}$$

are predicted as Forest, corresponding to 73.04%, while

$$888 \tag{6.96}$$

are predicted as Grassland, corresponding to 23.90%.
Together,

$$3602/3716 = 96.93\% \tag{6.97}$$

of the true Shrubland pixels are absorbed by Forest or Grassland.

This concentration is scientifically important. The class is not being distributed uniformly across unrelated classes. It is being absorbed almost entirely into neighbouring vegetation categories.

6.6.8 Clean-model confusion matrix

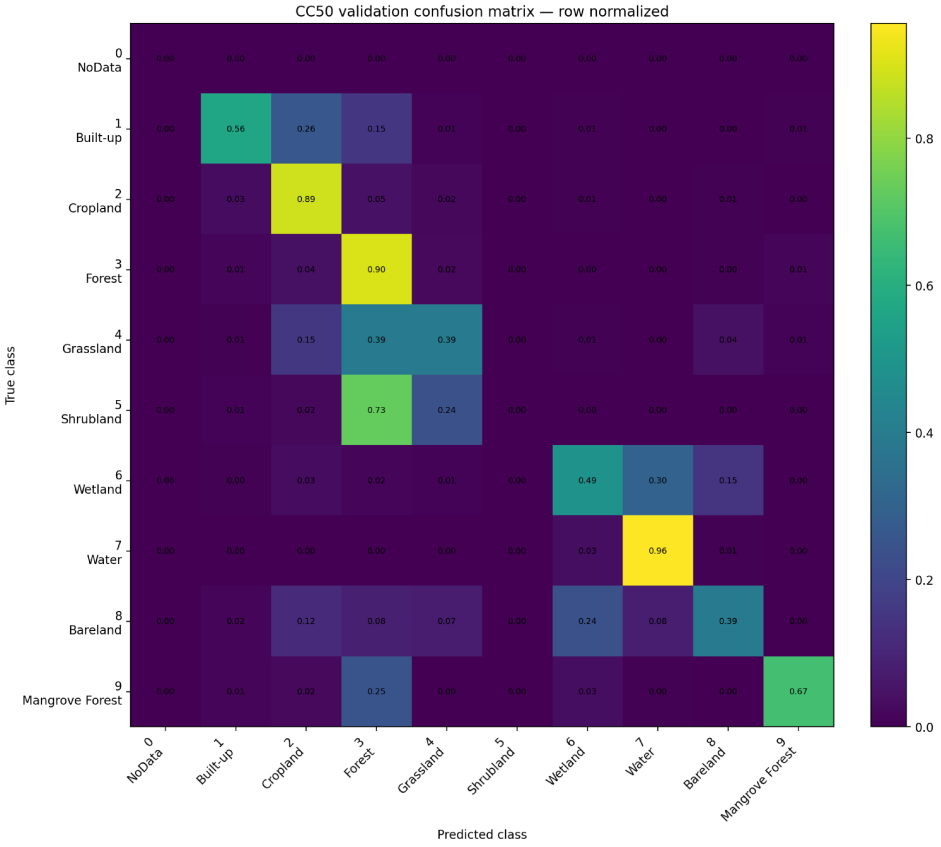


Figure 6.8: Row-normalized confusion matrix for the final CC50 clean baseline. Rows represent the MOLCA-derived reference classes, and columns represent the SwinUNETR predicted classes. NoData is excluded from the principal accuracy interpretation. The matrix reveals strong diagonal performance for Water, Forest, and Cropland, while highlighting the systematic misclassification of minority classes, particularly the complete absorption of Shrubland into neighbouring vegetation categories.

The matrix shows a strong diagonal for Water, Forest, and Cropland and a much weaker diagonal for the minority and transition classes.

The most prominent systematic patterns include:

- Built-up → Cropland and Forest;
- Grassland → Forest;
- Shrubland → Forest and Grassland;
- Wetland → Water and Bareland;
- Bareland → Wetland and Cropland;
- Mangrove Forest → Forest.

These patterns motivate the detailed error analysis in Section 6.7.

6.6.9 Effect of the 10× Shrubland sampler

The 10× sampler succeeds in changing the model's behaviour: the model no longer completely avoids class 5.

It produces:

$$26\,056 \tag{6.98}$$

Shrubland predictions.
However, only

$$39 \tag{6.99}$$

are true positives.
Therefore:

$$FP = 26\,017, \tag{6.100}$$

$$FN = 3677, \tag{6.101}$$

$$\text{precision} = 0.001497, \tag{6.102}$$

$$\text{recall} = 0.010495, \tag{6.103}$$

$$F_1 = 0.002620, \tag{6.104}$$

and

$$\text{IoU} = 0.001312. \tag{6.105}$$

Thus, increasing training-patch exposure from approximately 13.2% to 60.3% produces only approximately 1.05% Shrubland recall.

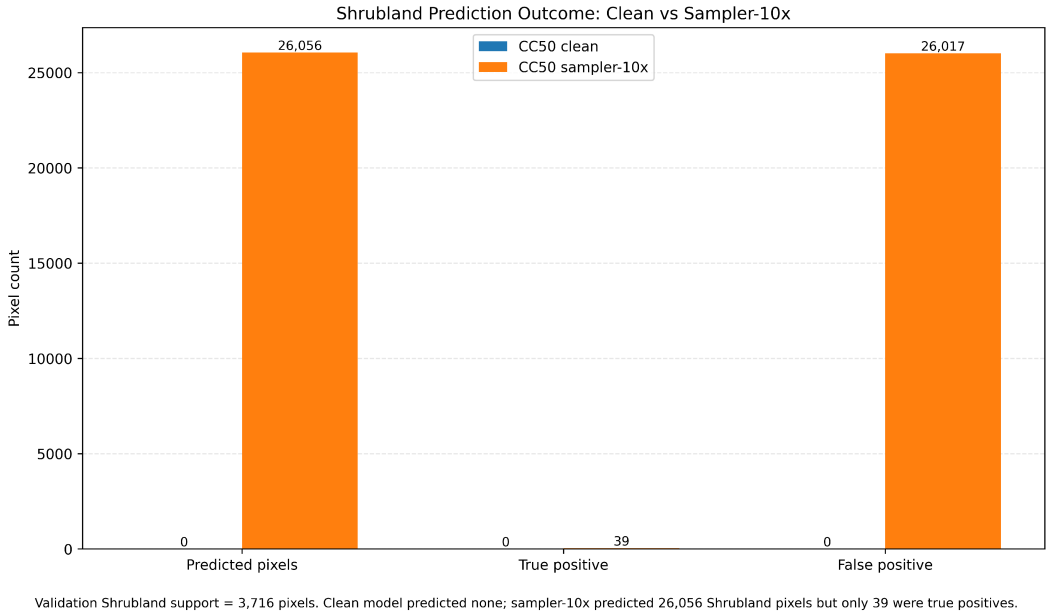


Figure 6.9: Shrubland prediction outcomes comparing the CC50 clean baseline and the 10× sampler experiment. The chart highlights that while the sampler successfully forces the model to predict class 5 (over 26,000 pixels), the overwhelming majority of these predictions are false positives, with only 39 true positives recovered.

6.6.10 Origin of the sampler's Shrubland predictions

The predicted class-5 pixels mainly originate from Forest reference pixels.

Of the 26,056 pixels predicted as Shrubland:

$$20\,457 \tag{6.106}$$

are actually Forest, corresponding to

$$78.51\%. \tag{6.107}$$

A further

$$2698 \tag{6.108}$$

are Grassland, corresponding to 10.36%.

Only 39 predicted class-5 pixels, or approximately 0.15%, correspond to actual Shrubland reference pixels.

Table 6.10: Reference-class origin of the 26,056 pixels predicted as Shrubland by the 10× sampler model.

Reference class	**Pixels predicted as Shrubland**	**Fraction**
Built-up	279	1.07%
Cropland	1,844	7.08%
Forest	20,457	78.51%
Grassland	2,698	10.36%
Shrubland	39	0.15%
Wetland	83	0.32%
Bareland	510	1.96%
Mangrove Forest	146	0.56%

The sampler therefore does not simply recover previously suppressed Shrubland. Instead, it shifts a portion of the vegetation decision space toward class 5, particularly at the expense of Forest.

6.6.11 Class-wise clean-versus-sampler changes

The sampler affects several classes beyond Shrubland.

Cropland recall decreases from

$$0.8869 \tag{6.109}$$

to

$$0.7666, \tag{6.110}$$

although its precision increases.

Grassland recall increases slightly from 0.3923 to 0.4280, but precision decreases substantially from 0.6154 to 0.4643.

Bareland recall increases from 0.3888 to 0.5006, but its precision falls from 0.4960 to 0.4039.

Mangrove performance deteriorates substantially:

$$\text{IoU} : 0.1553 \rightarrow 0.0804. \tag{6.111}$$

Wetland IoU also decreases:

$$0.3215 \rightarrow 0.2689. \tag{6.112}$$

The sampler therefore changes the multi-class decision boundaries more widely than class 5 alone.

6.6.12 Sampler confusion matrix

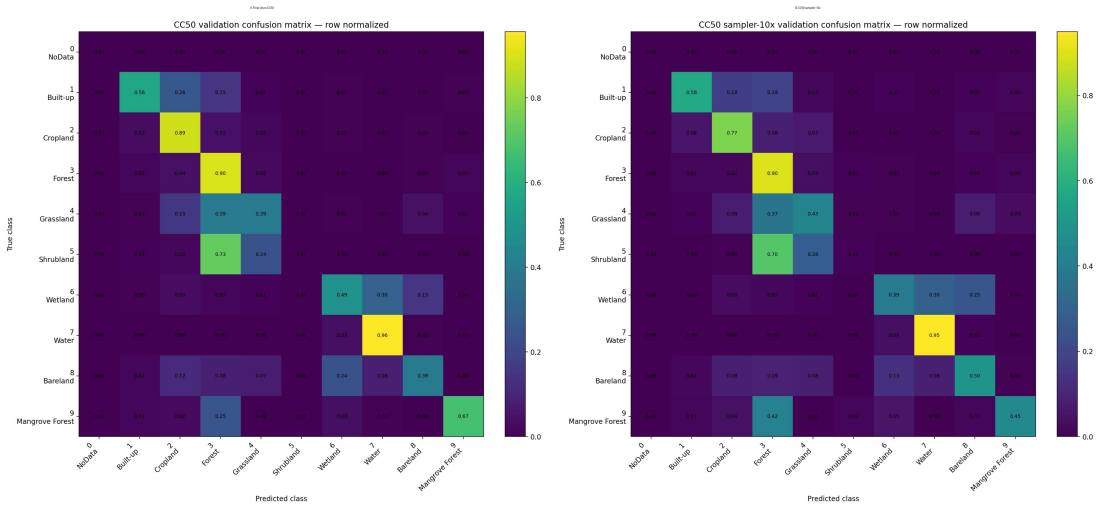


Figure 6.10: Side-by-side comparison of the row-normalized confusion matrices for the CC50 clean baseline (left) and the 10× WeightedRandomSampler experiment (right). The visual comparison confirms that heavily sampling Shrubland patches does not create a distinct Shrubland diagonal; instead, it redistributes false predictions across neighbouring vegetation categories.

The clean and sampler matrices together show that increased class-5 exposure does not create a clear Shrubland diagonal. Instead, a small diagonal value of approximately 0.0105 appears together with increased false class-5 assignments from several other classes.

6.7 Discussion of Errors and Limitations

6.7.1 What the final model learns reliably

The strongest result of the final CC50 clean baseline is the reliable segmentation of the dominant broad land-cover classes.

Water is the strongest class with IoU 0.8934 and recall 0.9561. The low backscatter behaviour of open water and its spatial coherence provide a relatively distinctive signal in Sentinel-1 data.

Forest reaches IoU 0.8102 and recall 0.9025. Sentinel-1 C-band time series have been shown to provide meaningful forest/non-forest discrimination, although achievable separability varies geographically [44]. The result here confirms that Forest can be represented effectively at the broad class level in the Philippine dataset.

Cropland reaches IoU 0.7607 and recall 0.8869. Its performance indicates that the temporal grouping and spatial statistical features provide substantial information on agricultural surfaces.

These three classes account for a large share of the valid validation pixels. Their strong performance explains the high OA, but the macro statistics and class-wise metrics demonstrate that they do not represent the complete difficulty of the ten-class problem.

6.7.2 Class imbalance matters, but does not explain all errors

Class imbalance is an important part of the problem. Remote-sensing semantic segmentation commonly exhibits severe differences in class frequency, and the literature identifies weighting, data augmentation, sampling, and model design as complementary strategies for addressing the resulting learning bias [45].

The weighted cross-entropy used in this thesis is consistent with published guidance for imbalanced thematic classification [46].

Nevertheless, the Shrubland experiment demonstrates why imbalance correction should not automatically be expected to solve a class-separability problem.

The 10× sampler increases the expected frequency of Shrubland-containing training patches from approximately 13.2% to 60.3%. This is a major change in training exposure.

Yet recall increases only from:

$$0 \tag{6.113}$$

to:

$$0.0105. \tag{6.114}$$

At the same time, 26,017 false Shrubland predictions are introduced and all principal global metrics decrease.

The experiment therefore provides direct project-specific evidence that limited patch exposure is not the only constraint.

6.7.3 Why the five Shrubland validation scenes are informative

The five validation scenes containing Shrubland are:

1. `51QTU_03_11`: 3191 model-space Shrubland pixels;
2. `50PRC_11_05`: 298;
3. `51PVN_06_05`: 115;
4. `51NVJ_09_04`: 67;
5. `51QTV_07_11`: 45.

The first scene alone contains approximately.

$$85.87\% \tag{6.115}$$

of all validation Shrubland pixels.

It is also the only scene in which the sampler obtains any Shrubland true positives.

This concentration means that a global Shrubland metric is heavily influenced by one geographic area. It also provides a useful natural diagnostic: the one scene with successful class-5 predictions can be compared with four scenes where class 5 remains entirely unsuccessful.

6.7.4 Patch-level Shrubland exposure

The amount of Shrubland contained in an individual 256×256 model patch varies substantially.

All nine overlapping patches in scene `51QTU_03_11` contain Shrubland. The strongest patch contains

$$1676 \tag{6.116}$$

Shrubland pixels, corresponding to approximately

$$2.56\% \tag{6.117}$$

of that patch.

The strongest fractions in the four remaining scenes are much smaller:

Table 6.11: Maximum Shrubland support within one 256×256 patch for the five Shrubland validation scenes.

Scene	Maximum Shrubland pixels	Maximum patch fraction
51QTU_03_11	1,676	2.557%
50PRC_11_05	298	0.455%
51PVN_06_05	115	0.175%
51NVJ_09_04	67	0.102%
51QTV_07_11	45	0.069%

This result supports the interpretation that local class exposure contributes to the difficulty. However, it is not sufficient by itself to explain the complete pattern, because substantial rescued Shrubland supervision is present in the training set.

6.7.5 Shrubland spatial geometry

An unexpected result is that the one validation scene producing true positives is not the most spatially coherent Shrubland example.

In model-space, `51QTU_03_11` contains 3191 Shrubland pixels divided across approximately 1045 connected components. The largest component contains only 56 pixels, and the median component size is one pixel.

By contrast:

- `50PRC_11_05` contains only two components;
- `51PVN_06_05` contains two components;
- `51NVJ_09_04` contains one component;
- `51QTV_07_11` contains one component.

Thus, the sampler's limited success in 51QTU cannot be explained by that scene containing the cleanest or largest coherent Shrubland objects.

6.7.6 Reference provenance of the validation Shrubland

The provenance analysis revealed a striking difference between the five Shrubland validation scenes.

At the native 549×549 reference resolution:

$$3718/3718 \tag{6.118}$$

Shrubland pixels in `51QTU_03_11` were inherited from the original MOLCA map.

The other four scenes were entirely rescue-derived:

$$332/332 \tag{6.119}$$

for `50PRC_11_05`,

$$86/86 \tag{6.120}$$

for `51NVJ_09_04`,

$$134/134 \tag{6.121}$$

for `51PVN_06_05`, and

$$57/57 \tag{6.122}$$

for `51QTV_07_11`.

All five patch references matched the aligned CC50 full-tile product with zero mismatch pixels.

Therefore, all 39 sampler true positives necessarily belong to original-MOLCA Shrubland.

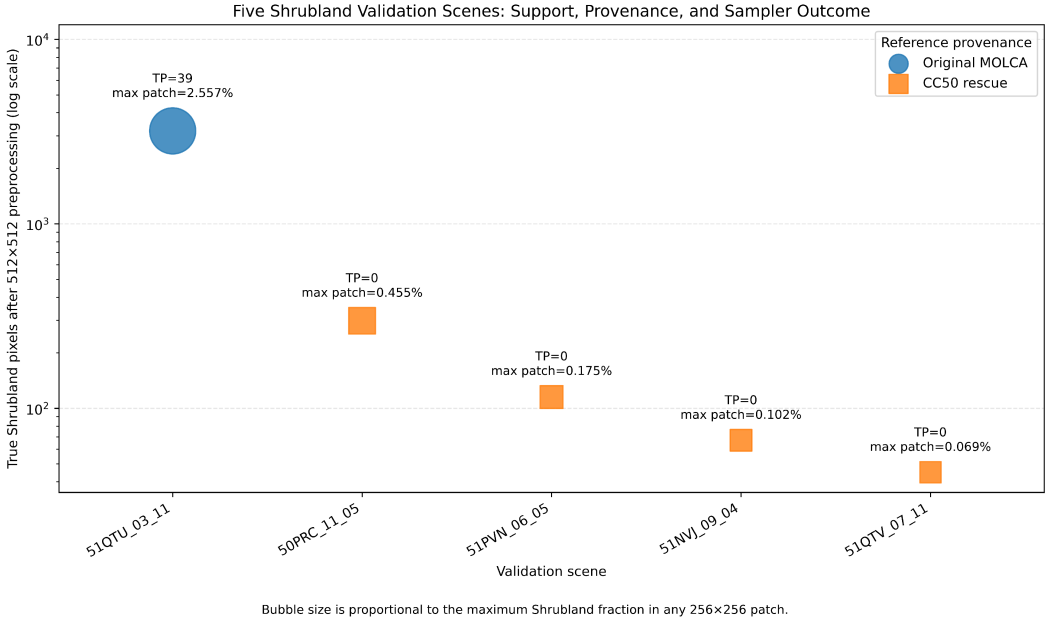


Figure 6.11: Analysis of the five validation scenes containing Shrubland, combining pixel support, reference provenance, maximum within-patch Shrubland fraction, and the 10× sampler true-positive outcome. The visualisation clearly demonstrates that the only scene producing true positives (51QTU_03_11) relies entirely on original MOLCA labels, whereas the four validation scenes derived exclusively from the CC50 rescue yield zero true positives.

This result is important but must be interpreted cautiously. It does *not* demonstrate that the CC50-rescued labels are incorrect. The training-set provenance analysis shows that the network received substantial rescue-derived Shrubland supervision during training. The observation instead motivated a deeper examination of whether Shrubland has a consistent SAR feature representation across the different geographic areas.

6.7.7 Posterior probabilities on true Shrubland

The final diagnostic moved beyond the hard argmax class and examined the model's posterior probabilities on the true Shrubland pixels.

For the clean model, the average Shrubland probability is low in every Shrubland validation scene.

Table 6.12: Mean clean-model probabilities on true Shrubland pixels.

Scene	**P(Forest)**	**P(Grassland)**	**P(Shrubland)**
51QTU_03_11	0.5295	0.2824	0.0453
50PRC_11_05	0.6446	0.2756	0.0249
51NVJ_09_04	0.9400	0.0321	0.0113
51PVN_06_05	0.7019	0.0568	0.0172
51QTV_07_11	0.8321	0.0449	0.0275

The result confirms that class 5 is not narrowly losing an otherwise close argmax competition in most scenes. The model assigns Forest substantially higher mean probability than Shrubland.

6.7.8 Sampler posterior behaviour

The sampler changes the final predictions but does not systematically increase probability on genuine Shrubland pixels.

The mean class-5 probabilities become approximately:

$$\begin{aligned} 51QTU: &\quad 0.0453 \to 0.0300, \\ 50PRC: &\quad 0.0249 \to 0.0078, \\ 51NVJ: &\quad 0.0113 \to 0.0036, \\ 51PVN: &\quad 0.0172 \to 0.0198, \\ 51QTV: &\quad 0.0275 \to 0.0142. \end{aligned} \tag{6.123}$$

Only `51PVN_06_05` shows a small increase in average Shrubland probability, yet it still produces no true class-5 argmax pixels.

Across the four rescue-derived validation scenes, the weighted mean Shrubland probability decreases from approximately

$$0.0217 \tag{6.124}$$

in the clean model to

$$0.0104 \tag{6.125}$$

in the sampler model.

This strengthens the interpretation that the sampler's 39 true positives do not represent a broad improvement in Shrubland confidence. Instead, the modified training distribution changes the decision boundary in a way that creates many more class-5 predictions overall.

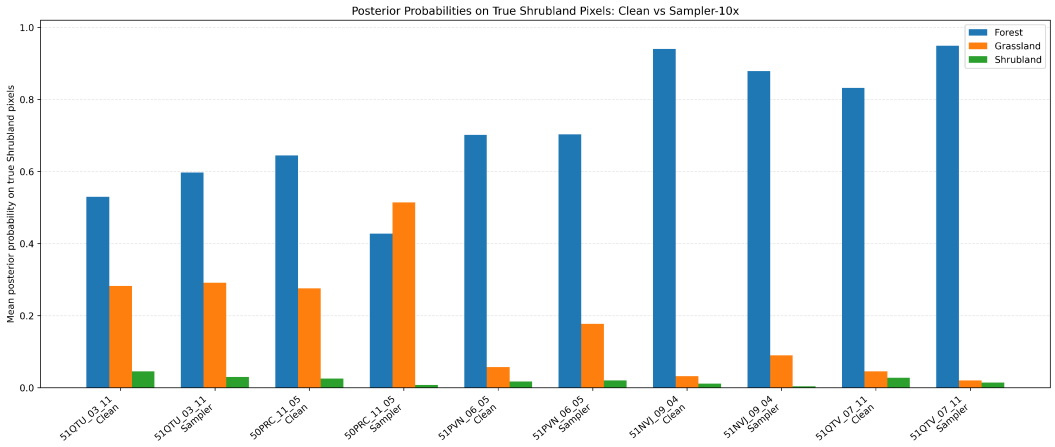


Figure 6.12: Mean posterior probabilities assigned to true Shrubland pixels by the clean CC50 model and the 10× sampler experiment. The diagnostic confirms that genuine Shrubland generally receives very low class-5 posterior probability, with the Forest class dominating the predictions in most scenes. Furthermore, the 10× sampler fails to systematically increase confidence in the correct class.

6.7.9 Cross-scene Shrubland behaviour

The five validation scenes containing Shrubland also reveal substantial variation in model behaviour across geographic areas. The largest scene, `51QTU_03_11`, contains 3191 of the 3716 validation Shrubland pixels and is the only scene in which the 10× sampler produces any correct Shrubland predictions. The other four scenes contain much smaller class-5 supports and remain at zero true positives.

This pattern is consistent with the posterior-probability analysis. In the clean model, the mean Shrubland posterior remains low in every one of the five scenes, while Forest receives the highest mean posterior on true Shrubland pixels. The sampler changes the final decision boundary sufficiently to create class-5 predictions, but this change is not reproduced consistently across the five geographic areas.

The result should therefore be interpreted at the model-output level: the complete 28-channel Sentinel-1 representation supports strong mapping of several major classes, but the learned Shrubland decision is not stable across the examined validation scenes. This demonstrates that class-5 generalisation remains geographically uneven in the final model.

6.7.10 Class imbalance versus cross-scene generalisation

Taken together, the project evidence supports a more specific interpretation than class imbalance alone.

Class imbalance clearly contributes to the Shrubland problem:

- Shrubland is globally rare;
- only five validation scenes contain the class;
- several validation patches contain very small Shrubland fractions.

However, imbalance alone does not adequately explain the final behaviour. The strongest evidence is:

1. the 10× sampler increases expected Shrubland-patch exposure from approximately 13.2% to 60.3%, yet recall reaches only 1.05%;
2. 85.16% of the Shrubland training pixels already come from the CC50 rescue;
3. the four validation scenes containing only rescued Shrubland still produce zero true positives;
4. mean Shrubland posterior probabilities remain low in every examined scene;
5. all 39 sampler true positives occur in a single validation scene, while the remaining four scenes produce none.

The accumulated evidence therefore points to a combined rare-class, reference-representation, and cross-scene generalisation problem rather than to insufficient training frequency alone. Increasing patch exposure shifts some vegetation predictions toward Shrubland, but it does not create a stable class-5 response across the validation geography.

6.7.11 Limitations associated with the reference data

The reference dataset remains an agreement-derived product rather than an independent field-survey ground truth.

This has several implications.

First, agreement between products increases confidence but does not guarantee perfect thematic correctness.

Second, the input products represent different nominal years. Temporal mismatch may therefore create local disagreement unrelated to pure classification error.

Third, the revised CC50 Shrubland labels are derived from a specific conservative rule involving NAMRIA, FROM-GLC, and previously unclassified MOLCA pixels. They should not be interpreted as an independently validated national Shrubland inventory.

Fourth, the connected-component threshold introduces a spatial prior: rescued Shrubland objects smaller than 50 connected pixels are discarded. This improves spatial coherence but may remove genuinely small Shrubland patches.

The CC50 reference is therefore best interpreted as a controlled, conservative training reference rather than as definitive national ground truth.

6.7.12 Limitations of VH-only Sentinel-1 features

The present model uses Sentinel-1 VH-derived features only.

This has operational advantages. The processing chain remains relatively compact and follows the design of the CCI+ workflow. SAR also provides important cloud-independent observations in a tropical environment.

However, a single polarisation does not contain all potentially available information.

The literature shows that multi-temporal Sentinel-1 can discriminate broad forest/non-forest conditions, but the level of separability varies between geographic sites [44]. For more specific ecosystem classes, multi-source combinations can provide additional discrimination. Mangrove studies, for example, often benefit from combining optical and SAR information [43]. Complex wetland classification likewise commonly benefits from complementary Sentinel-1 and Sentinel-2 temporal features [42].

These studies do not imply that optical data or VV polarisation would automatically solve the present errors. They show that the information content available from VH-only SAR is only one subset of the possible feature space.

Any future addition of VV or optical information would constitute a new experiment and would need to be evaluated against the current controlled baseline.

6.7.13 No independent held-out test set

The 28 validation scenes were used both for model-selection monitoring and for the detailed final internal validation.

The final metrics are therefore not estimates from a completely untouched third test set.

The distinction is important because the best checkpoint is selected based on validation mIoU. The final evaluation subsequently provides a more complete reconstruction-based assessment of the same validation population.

The reported values should therefore be described as:

internal validation performance on the fixed 28-scene validation set

rather than as independent external test accuracy.

A stronger future evaluation would use geographically separate scenes not used for model selection or, ideally, an independently validated reference dataset.

6.7.14 Interpretation of the final model

The final CC50 clean model should be interpreted as a successful baseline for large-scale SAR-based Philippine land-cover segmentation, but not as a uniformly accurate classifier for every target class.

The model performs strongly on Water, Forest, and Cropland and moderately on. Built-up, Grassland, Wetland, and Bareland.

Mangrove Forest achieves good recall but low precision, indicating overprediction.

Shrubland remains unresolved.

The failed 10× sampler experiment is scientifically useful because it narrows the interpretation of the problem. The class is not recovered simply by presenting Shrubland-containing patches far more frequently.

The later reference-provenance and posterior diagnostics further show that:

- the revised reference substantially increased Shrubland supervision;
- most training Shrubland is rescue-derived;
- Forest dominates the Shrubland posterior in most validation areas;
- the sampler's limited success is concentrated in only one of the five Shrubland validation scenes;

The most defensible conclusion is therefore not that one single component of the pipeline is responsible for the Shrubland failure. Instead, the result appears to emerge from the interaction of rare-class frequency, reference construction, geographic heterogeneity, and uneven generalisation of the learned class boundary across the available validation scenes.

This interpretation is consistent with broader remote-sensing research in which both class imbalance and geographic/domain variability limit the transferability of semantic segmentation models [45, 47, 48].

6.7.15 Implications for further experimentation

The evidence suggests that another arbitrary increase in sampler strength would not be the most informative immediate experiment.

A more scientifically controlled sequence would be:

1. independently inspect and validate Shrubland labels in a targeted subset of original-MOLCA and CC50-rescued scenes;
2. perform controlled feature-ablation experiments to quantify the contribution of the different Sentinel-1 feature families;
3. evaluate whether additional Sentinel-1 information, particularly VV polarisation, improves separation between Forest, Grassland, and Shrubland;
4. evaluate geographically balanced or geographically disjoint training and validation strategies;
5. only after these checks, reconsider more specialised rare-class loss functions or sampling strategies;
6. if the application requirements permit, investigate multimodal Sentinel-1–Sentinel-2 inputs for the most difficult land-cover classes.

The current clean CC50 model should remain the reference point for these future controlled experiments.

6.8 Chapter Summary

This chapter presented the results of the complete MOLCA–Sentinel-1– Swin-UNETR workflow for Philippine land-cover mapping.

The original 60-tile MOLCA reference contained 5.80 billion valid labelled pixels but only 16,699 Shrubland pixels. A later source-to-final audit showed that Shrubland existed in the input products but was largely lost during the agreement hierarchy. A conservative Shrubland rescue was therefore applied only where the original reference was unclassified, and NAMRIA and FROM-GLC both indicated Shrubland.

The raw rescue recovered approximately 5.59 million pixels but was highly fragmented. Connected-component filtering showed a clear trade-off between class enrichment and spatial coherence. CC50 retained 783,561 rescued pixels and produced 800,260 total Shrubland pixels. It was selected as the principal revised reference after quantitative and visual comparison with CC10 and CC100.

The final deep learning dataset contained 277 valid scenes and 7756 Sentinel-1 feature rasters. The CC50 training set contained 62,501 model-space Shrubland pixels, 85.16% of which originated from the conservative rescue.

The final clean CC50 SwinUNETR model stopped at epoch 78, with its best checkpoint at epoch 53. Full-scene internal validation produced an Overall Accuracy of 0.8333, macro F1 of 0.5460, mean IoU of 0.4339, and Cohen's Kappa of 0.7703.

Water, Forest, and Cropland were the strongest classes, with IoU values of 0.8934, 0.8102, and 0.7607, respectively. Built-up, Grassland, Wetland, and Bareland showed moderate performance. Mangrove Forest achieved relatively high recall but low precision. Shrubland remained completely undetected by the clean model.

A controlled sampler with a 10× weight for Shrubland-containing patches increased their expected training exposure from approximately 13.2% to 60.3%. The sampler produced 26,056 Shrubland predictions, but only 39 were correct. Shrubland recall reached only 0.0105 while more than 26,000 false positives were introduced. All principal global metrics decreased. The sampler was therefore not considered an effective solution.

Further diagnosis showed that all 39 true positives occurred in one of the five Shrubland validation scenes. Most training Shrubland supervision already came from the CC50 rescue, so failure on the rescued validation areas could not be explained simply by the absence of rescued labels during training. Posterior analysis showed persistently low class-5 probabilities and strong scene-to-scene variation in the model's ability to recover Shrubland.

Overall, the results demonstrate that Sentinel-1 VH seasonal features and a SwinUNETR model can provide effective mapping of several major Philippine land-cover classes. At the same time, the rare vegetation classes reveal the importance of reference-map design, spatially representative training data, class separability, and geographic generalisation in large-area semantic segmentation.

Chapter 7

Conclusions

This thesis investigated the development of a deep learning workflow for high-resolution land-cover mapping in the Philippines using Sentinel-1 Synthetic Aperture Radar (SAR) data and a training reference derived from the Map of Land Cover Agreement (MOLCA) methodology.

The work addressed several connected challenges. The first was the creation of a sufficiently reliable reference dataset over a large and geographically heterogeneous tropical country. The second was the construction of a representative patch dataset from the national reference. The third was the generation and organisation of multi-temporal Sentinel-1 features suitable for semantic segmentation. The final challenge was the adaptation, training, evaluation, and analysis of a transformer-based segmentation model under a strongly imbalanced multi-class land-cover setting.

Rather than treating these steps independently, the thesis developed them as parts of a single processing chain. This proved particularly important when the first full deep learning experiment exposed a complete failure for the Shrubland class. Investigation of this result led back to the reference dataset, showing that the scarcity of Shrubland was not simply a neural network problem but was also related to the way agreement-based labels had been generated. A conservative reference revision was therefore developed, evaluated, and used to construct the final CC50 training dataset.

The final clean CC50 model achieved strong internal validation performance for several dominant land-cover classes while also revealing limitations of the current SAR-only approach for rare and structurally similar vegetation classes. These results provide a reproducible internal baseline for the Philippine study area and a basis for further methodological development.

7.1 Summary

7.1.1 Overview of the developed workflow

The workflow developed in this thesis begins with the construction of the Philippine MOLCA reference and ends with a trained semantic segmentation model evaluated over geographically distributed validation scenes.

The main stages were:

1. identification and preparation of global and regional land-cover products covering the Philippines;
2. harmonisation of the different source legends into a common ten-code reference legend comprising NoData and nine land-cover classes;
3. spatial alignment of the source products to a common reference grid;
4. generation of a MOLCA-derived agreement map using general land-cover products, thematic evidence products, and regional Philippine/Southeast Asian information;
5. selection of 60 Sentinel-2 tiles covering representative areas of the Philippines;
6. generation and analysis of candidate 549×549 reference patches;
7. selection of a geographically and thematically diverse subset of 280 patches;
8. collection and preparation of 2019 Sentinel-1 VH observations;
9. generation of four temporal groups and seven SAR feature products per group, resulting in 28 model-input features;
10. association and verification of the Sentinel-1 feature stacks with their corresponding MOLCA reference patches;
11. adaptation and training of a SwinUNETR-based semantic segmentation model;
12. evaluation using class-specific and global accuracy metrics;
13. investigation of rare-class errors, with particular attention to Shrubland;
14. revision of the Shrubland reference using a conservative agreement-based rescue strategy;

15. controlled comparison between the final clean CC50 model and a Shrubland-focused weighted-sampling experiment.

This sequence allowed the study to address the complete path from heterogeneous existing land-cover products to a trained SAR semantic segmentation model rather than considering only the final classification stage.

7.1.2 Philippine adaptation of the MOLCA concept

The MOLCA methodology provides an alternative to producing a large training reference entirely from manual interpretation. Instead, multiple existing land-cover datasets are brought into a common spatial and thematic framework, and areas of sufficiently strong agreement are retained as training information.

The Philippine implementation required substantial adaptation because the available datasets differed in spatial resolution, coordinate reference system, nominal year, class nomenclature, geographic coverage, and thematic purpose.

The final workflow incorporated eight principal input products. FROM-GLC and GlobeLand30 provided the main general land-cover information. WSF, GHS-BUILT-S, GSW, and JAXA FNF provided thematic evidence for Built-up, Water/Wetland, and Forest-related classes. NAMRIA supplied national Philippine land-cover information, while the JAXA high-resolution Southeast Asia product provided additional regional information.

The use of regional products was particularly important for the Philippine context. It enabled the preservation of Mangrove Forest as an explicit land-cover class and allowed regionally relevant categories to contribute to areas that were unresolved by the global products.

The resulting reference used nine valid land-cover classes:

1. Built-up;
2. Cropland;
3. Forest;
4. Grassland;
5. Shrubland;
6. Wetland;
7. Water;
8. Bareland;

9. Mangrove Forest,

with NoData/unclassified pixels handled separately.

The work therefore demonstrates that an agreement-based reference concept can be transferred to the Philippines, but that the transfer requires careful adaptation of the source products and class logic rather than direct application of a pre-existing workflow.

7.1.3 Patch and Sentinel-1 dataset construction

The study initially selected 280 geographical patches distributed across 60 Sentinel-2 tiles.

The purpose of this selection was not to reproduce the national class distribution proportionally. Instead, the patches were selected to obtain broad geographical coverage while also retaining less common and classification-relevant land-cover conditions.

During the Sentinel-1 verification stage, three patches belonging to one tile could not provide the complete required 2019 SAR input. These samples were excluded rather than artificially completing the missing data.

The final deep learning dataset therefore contained:

$$277 \tag{7.1}$$

valid scenes, divided into:

$$249 \tag{7.2}$$

training scenes and

$$28 \tag{7.3}$$

validation scenes.

For every valid scene, 28 Sentinel-1 features were associated with the corresponding reference map.

The feature representation consisted of four temporal groups, each containing.

$$\{\text{SI}, \text{LEE}, \text{MAX}, \text{MIN}, \text{MAXMIN}, \text{MEAN}, \text{MEDIAN}\}. \tag{7.4}$$

Thus,

$$4 \times 7 = 28 \tag{7.5}$$

features were supplied to the segmentation model.

The final valid dataset therefore contained

$$277 \times 28 = 7756 \tag{7.6}$$

Sentinel-1 feature rasters.

The work confirms that careful control of input/reference association, dimensions, coordinate reference systems, feature ordering, and train–validation separation is essential before model performance can be interpreted.

7.1.4 Development from the original reference to CC50

One of the most important outcomes of the thesis was the discovery that the original MOLCA reference contained an extremely small amount of Shrubland.

Across the complete 60-tile reference, the original final MOLCA product contained only

$$16\,699 \tag{7.7}$$

Shrubland pixels.

This was not sufficient to represent the class adequately relative to dominant categories such as Forest, Cropland, and Water.

The problem became particularly evident after the first complete training experiment, in which Shrubland was not detected by the model.

A source-to-final audit subsequently showed that potentially useful Shrubland information existed in the input products, particularly where NAMRIA and FROM-GLC independently identified the same class.

A conservative rescue rule was therefore introduced:

$$\begin{aligned} &\text{Original MOLCA} = \text{unclassified} \\ &\wedge\ \text{NAMRIA} = \text{Shrubland} \\ &\wedge\ \text{FROM-GLC} = \text{Shrubland} \\ &\Rightarrow \text{Shrubland}. \end{aligned} \tag{7.8}$$

The conservative formulation was important because already classified Forest, Grassland, Cropland, or other valid pixels were not replaced.

The initial rescue recovered approximately 5.59 million candidate Shrubland pixels. However, connected-component analysis demonstrated that the raw rescue contained a large amount of spatial fragmentation.

For this reason, CC10, CC50, and CC100 variants were generated.

CC10 retained more Shrubland but remained relatively fragmented. CC100 was spatially conservative but removed a large fraction of the recovered information. CC50 represented an intermediate solution.

The final CC50 reference contained:

$$800\,260 \tag{7.9}$$

Shrubland pixels across the 60-tile reference, of which

$$783\,561 \tag{7.10}$$

were introduced through the component-filtered rescue.

CC50 was consequently adopted as the principal revised reference for the final controlled deep learning experiments.

7.2 Main Findings

7.2.1 Performance of the final clean CC50 model

The final reference experiment was the clean CC50 SwinUNETR model.

The experiment used the fixed 249/28 train–validation split and retained the same general model configuration used for the controlled comparison.

Training stopped at epoch 78 through the early-stopping criterion.

The best checkpoint was obtained at:

$$E_{\text{best}} = 53, \tag{7.11}$$

with a patch-level validation mIoU of:

$$0.3697. \tag{7.12}$$

The final reconstructed full-scene evaluation was performed over

$$6\,174\,226 \tag{7.13}$$

valid validation pixels.

The main global results were:

$$\text{OA} = 0.8333, \tag{7.14}$$

$$F_{1,\text{macro}} = 0.5460, \tag{7.15}$$

$$\text{mIoU} = 0.4339, \tag{7.16}$$

and

$$\kappa = 0.7703. \tag{7.17}$$

These results show that the developed Sentinel-1/SwinUNETR workflow learned a useful land-cover representation across the fixed validation sample. Because the same validation population was used for model monitoring and final reconstructed assessment, the results should be interpreted as internal validation performance rather than as evidence of national-scale transferability.

The performance should nevertheless be understood at class level rather than through Overall Accuracy alone.

7.2.2 Strongest land-cover classes

Water was the strongest class in the final clean model.

Its IoU reached:

$$0.8934, \tag{7.18}$$

with recall:

$$0.9561. \tag{7.19}$$

Forest was also classified strongly:

$$\mathrm{IoU}_{\mathrm{Forest}} = 0.8102, \tag{7.20}$$

and

$$R_{\mathrm{Forest}} = 0.9025. \tag{7.21}$$

Cropland achieved:

$$\mathrm{IoU}_{\mathrm{Cropland}} = 0.7607 \tag{7.22}$$

and recall:

$$0.8869. \tag{7.23}$$

These three classes represent the strongest part of the final internal validation performance.

The use of seasonal Sentinel-1 features is consistent with the broader spatio-temporal SAR mapping approach described by Russo et al. [9], where seasonal synthesis provides useful temporal information without requiring every individual acquisition to be supplied directly to the model.

7.2.3 Intermediate-performance classes

Built-up, Grassland, Wetland, and Bareland achieved intermediate performance.

Built-up obtained an IoU of approximately 0.3708.

Grassland obtained an IoU of 0.3150, with a relatively important confusion toward Forest.

Wetland obtained an IoU of 0.3215 and was frequently confused with Water.

Bareland obtained an IoU of 0.2787 and showed a more distributed confusion pattern involving Wetland, Cropland, Forest, Water, and Grassland.

These results demonstrate that classes with transitional physical characteristics or strong local heterogeneity are more difficult to distinguish than the dominant broad classes.

7.2.4 Mangrove Forest result

Mangrove Forest produced a different type of error.

Its recall reached approximately:

$$0.6717, \tag{7.24}$$

which indicates that a substantial fraction of the reference Mangrove pixels was detected.

However, precision was only:

$$0.1680. \tag{7.25}$$

The model therefore substantially overpredicted Mangrove.

This result is important because it demonstrates that a minority class can fail in different ways.

Shrubland was mainly under-detected, whereas Mangrove Forest was detected with relatively high recall but poor precision.

Consequently, rare-class performance cannot be addressed using only one generic balancing strategy.

7.2.5 Shrubland remained the principal unresolved class

The most important unresolved classification problem is Shrubland.

The CC50 reference substantially increased Shrubland supervision.

Within the final training data, the model-space training set contained:

$$62\,501 \tag{7.26}$$

Shrubland pixels distributed across 60 scenes.

The validation set contained:

$$3716 \tag{7.27}$$

Shrubland pixels across five scenes.
Nevertheless, the final clean CC50 model predicted no Shrubland pixels.
Therefore:

$$TP = 0, \tag{7.28}$$

$$FN = 3716, \tag{7.29}$$

and:

$$\mathrm{IoU}_{\mathrm{Shrubland}} = 0. \tag{7.30}$$

The confusion pattern was highly structured.
Approximately:

$$73.0\% \tag{7.31}$$

of true Shrubland pixels were classified as Forest, while approximately:

$$23.9\% \tag{7.32}$$

were classified as Grassland.
Together, Forest and Grassland therefore absorbed approximately:

$$96.9\% \tag{7.33}$$

of the validation Shrubland.

This result suggests that the principal difficulty is related to separation within the vegetation classes rather than random confusion throughout the legend.

7.2.6 The 10× sampler did not solve the problem

A controlled WeightedRandomSampler experiment was subsequently used to test whether much stronger exposure to Shrubland-containing patches could improve the class.

Training patches containing Shrubland received a sampling weight of 10, whereas the remaining patches received a weight of 1.

This increased the expected fraction of Shrubland-containing patch draws from approximately:

$$13.2\% \tag{7.34}$$

to:

$$60.3\%. \tag{7.35}$$

The sampler therefore represented a substantial increase in rare-class exposure. However, its final global performance decreased:

$$\mathrm{OA} : 0.8333 \rightarrow 0.8038, \tag{7.36}$$

$$\mathrm{mIoU} : 0.4339 \rightarrow 0.4038, \tag{7.37}$$

$$F_{1,\mathrm{macro}} : 0.5460 \rightarrow 0.5122, \tag{7.38}$$

and

$$\kappa : 0.7703 \rightarrow 0.7323. \tag{7.39}$$

The sampler produced:

$$26\,056 \tag{7.40}$$

Shrubland predictions.
Only:

$$39 \tag{7.41}$$

were true positives.
This resulted in:

$$\mathrm{Precision}_{\mathrm{Shrubland}} \approx 0.0015, \tag{7.42}$$

$$\mathrm{Recall}_{\mathrm{Shrubland}} \approx 0.0105, \tag{7.43}$$

and

$$\mathrm{IoU}_{\mathrm{Shrubland}} \approx 0.0013. \tag{7.44}$$

The majority of false Shrubland predictions originated from Forest.

The experiment therefore demonstrated that stronger sampling changed the frequency with which the network predicted class 5, but did not create a reliable Shrubland representation.

This result is consistent with the wider remote-sensing literature, where class imbalance is recognised as an important problem but one that may require a combination of data, loss, sampling, and representation-level solutions rather than oversampling alone [45].

7.2.7 Rare-class difficulty is not explained by frequency alone

The detailed Shrubland analysis provided one of the most important interpretations of this thesis.

Of the 62,501 Shrubland training pixels:

$$53\,225 \tag{7.45}$$

or approximately

$$85.16\% \tag{7.46}$$

came from the CC50 rescue.

Thus, the network was not trained primarily on the small amount of original-MOLCA Shrubland.

The model had substantial exposure to newly rescued Shrubland during training.

Nevertheless, the four validation scenes composed entirely of CC50-rescued Shrubland produced no correct class-5 predictions.

The one validation scene in which the sampler obtained 39 true positives was the scene whose Shrubland labels were inherited from the original MOLCA reference.

This observation alone does not prove that the rescued labels are wrong.

However, it showed that the remaining difficulty could not be reduced to the simple statement that the network had not seen enough rescued Shrubland.

7.2.8 Cross-scene variation in Shrubland behaviour

The five validation scenes containing Shrubland did not behave uniformly.

All 39 true-positive Shrubland pixels produced by the 10× sampler occurred in a single validation scene, while the other four Shrubland scenes still produced zero true positives.

The posterior analysis reported in Chapter 6 also showed that genuine Shrubland generally received low class-5 probability and was most often assigned to Forest or Grassland.

This scene-to-scene variation is important because it shows that the remaining Shrubland problem cannot be summarised by one national class-frequency value alone. The available evidence is consistent with an interaction between class imbalance, reference provenance, local vegetation characteristics, and geographic generalisation.

Accordingly, the present thesis does not claim that a single feature-space mechanism explains the Shrubland failure. A stronger conclusion would require additional geographically independent data and targeted validation of the Shrubland reference.

7.2.9 Overall methodological finding

The central methodological conclusion of this thesis is that reliable large-area land-cover mapping cannot be treated as a model-only optimisation problem.

The final classification behaviour depends on the interaction between:

- reference-map construction;
- source-product agreement;
- class definitions;
- geographic sampling;
- rare-class representation;
- SAR feature information;
- model architecture;
- loss weighting;
- training-patch sampling;
- geographic variability.

The Shrubland case demonstrates this particularly clearly.

The original problem first appeared as zero model recall.

However, investigation showed that the reference dataset itself contained almost no Shrubland.

After the reference was improved, the class still remained difficult.

Oversampling then showed that frequency alone was not sufficient.

Finally, the scene-level posterior and prediction analyses showed that Shrubland behaviour varied markedly across the five validation scenes.

The final result is therefore not simply a failed rare class. It provides a case study of how data quality, class representation, and geographic generalisation interact in large-area remote-sensing semantic segmentation.

7.3 Limitations

7.3.1 Agreement-derived reference rather than independent ground truth

The principal limitation of the study is that the training and validation reference is derived from agreement between existing land-cover products.

The MOLCA approach provides a scalable way to construct large reference datasets, but agreement between existing maps is not equivalent to independent field validation.

Two source products can agree while still sharing a classification error.

Conversely, two correct products can disagree because of differences in legend definitions, nominal year, spatial resolution, or minimum mapping unit.

The final reference should therefore be interpreted as a high-confidence agreement-derived training reference rather than as absolute ground truth.

7.3.2 Temporal differences between source products

The source land-cover products were not all produced for exactly the same year.

This was unavoidable because no single collection of global and regional high-resolution land-cover datasets exists for an identical observation period.

Consequently, part of the disagreement between source products may represent real land-cover change rather than thematic classification inconsistency.

This is particularly relevant in rapidly changing urban, agricultural, and coastal areas.

7.3.3 CC50 Shrubland remains an indirectly derived class

The CC50 revision increased Shrubland representation considerably, but the new labels were not independently field-validated.

They were obtained only where:

1. the original final MOLCA map was unclassified;
2. NAMRIA identified Shrubland;
3. FROM-GLC identified Shrubland;

4. the rescued component contained at least 50 connected pixels.

This is deliberately conservative, but it still represents a rule-based reference construction.

The connected-component threshold also introduces a minimum spatial-size constraint.

A genuine Shrubland patch smaller than the threshold may therefore be removed, while a sufficiently large but incorrect source agreement could remain.

CC50 should consequently be considered an improved training reference, not an independently validated Shrubland inventory.

7.3.4 Remaining class imbalance

Even after the CC50 revision, the class distribution remains highly imbalanced.

Water, Forest, and Cropland contain far more training pixels than Shrubland, Mangrove Forest, Wetland, and Bareland.

The use of class-weighted cross-entropy partly addresses this imbalance, but does not remove the underlying difference in spatial support, scene frequency, or geographic diversity.

The sampler experiment additionally demonstrated that increasing patch frequency alone is not sufficient.

7.3.5 Unequal geographic representation of rare classes

Rare classes are not uniformly distributed across the Philippines.

For Shrubland, only five validation scenes contain the class.

Furthermore, approximately 85.9% of all validation Shrubland pixels occur in a single scene.

This creates an important interpretative limitation.

A global Shrubland metric can be dominated by the behaviour of one geographic area and may not accurately describe national-scale class generalization.

7.3.6 SAR-only and VH-only input

The model uses Sentinel-1 SAR information and specifically the VH polarisation for the 28 seasonal features.

This was a deliberate design choice and maintains continuity with the spatio-temporal SAR methodology adapted in this work.

However, some land-cover distinctions may require information that is only weakly represented in VH backscatter.

Forest, Grassland, and Shrubland can share structural and moisture-related radar characteristics, while classes such as Mangrove Forest have ecological definitions that are partly related to geographic and hydrological context.

The current results therefore represent the capability of the selected SAR-only feature configuration rather than an upper limit on what could be achieved using additional sensor information.

7.3.7 Fixed model architecture and limited hyperparameter exploration

The purpose of the thesis was not to perform an exhaustive neural-network architecture search.

The SwinUNETR-based framework was selected because the work aimed to adapt and extend an existing Sentinel-1 deep learning methodology to the Philippines.

Consequently, the reported results do not establish that SwinUNETR is the optimal architecture for this dataset.

Alternative transformer models, CNN–Transformer hybrids, or more recent remote-sensing foundation-model approaches were outside the principal scope of the study.

7.3.8 No completely independent test set

The final 28 validation scenes were used both for validation monitoring during training and for the reconstructed final accuracy assessment.

The reported metrics are therefore internal validation results.

Although full-scene reconstruction prevents overlapping validation pixels from being counted repeatedly, it does not make the validation set an independent external test set.

A stronger future evaluation should use geographically separate data that have not contributed to checkpoint selection.

7.3.9 Internal validation and national interpretation

The Overall Accuracy of 0.8333 should not be interpreted as a statistically designed estimate of national Philippine land-cover map accuracy.

The 280 original patches were selected to support model training and improve geographic and thematic coverage, not through a probability sampling design intended for national accuracy estimation.

The reported accuracy therefore describes the performance of the model on the fixed project validation dataset.

7.4 Future Work

7.4.1 Independent validation of the MOLCA-derived reference

The most important future improvement is the creation of an independent validation dataset.

This could be based on:

- manually interpreted very-high-resolution imagery;
- authoritative national reference information;
- field observations where available;
- carefully reviewed stratified samples from multiple Philippine regions.

Such a dataset would allow the quality of the MOLCA-derived labels and the deep learning predictions to be evaluated independently.

Particular attention should be given to Shrubland, Mangrove Forest, Grassland, Wetland, and Bareland.

7.4.2 Targeted Shrubland reference validation

The next Shrubland-specific step should not simply be stronger oversampling.

Instead, a targeted reference study should compare three groups:

1. original-MOLCA Shrubland;
2. CC50-rescued Shrubland;
3. Forest/Grassland pixels surrounding the rescued regions.

A representative subset could be interpreted independently using high-resolution optical imagery and, where available, regional or national reference information.

This would help determine whether the observed cross-scene variability in Shrubland performance reflects:

- genuine diversity within the Shrubland class;
- residual reference inconsistency;
- differences in vegetation structure;
- environmental variability;
- or a combination of these factors.

7.4.3 Sensitivity analysis of the CC threshold

CC50 was selected after quantitative component analysis and visual comparison.

Future work could test this choice more directly by training otherwise identical models using:

$$\text{CC10}, \quad \text{CC50}, \quad \text{CC100}. \tag{7.47}$$

Such an experiment would determine whether the spatial-coherence threshold produces a measurable effect on model generalisation rather than only on the appearance of the reference map.

CC100 is particularly useful as a strict comparison because it retains only the larger and more spatially coherent rescue regions.

7.4.4 Dual-polarization Sentinel-1 features

A direct extension of the current SAR-only framework would be to include VV information in addition to VH.

VH and VV respond differently to vegetation structure, surface scattering, and moisture conditions.

A controlled dual-polarisation experiment could therefore test whether the additional information improves separation between:

$$\text{Forest}, \quad \text{Grassland}, \quad \text{Shrubland}. \tag{7.48}$$

The important requirement would be to change only the input representation while preserving the same split, reference, evaluation procedure, and model configuration as far as possible.

7.4.5 Additional SAR temporal information

The present approach compresses Sentinel-1 observations into four temporal groups and seven derived features per group.

This representation is computationally efficient and follows the spatio-temporal approach of Russo et al. [9].

However, future experiments could investigate whether difficult vegetation classes benefit from:

- finer temporal grouping;
- explicit monthly or phenological descriptors;
- seasonal variability measures;

- VV/VH relationships;
- interferometric coherence where acquisition geometry permits.

Such features could provide additional information about vegetation dynamics that is not fully represented by the current seasonal statistics.

7.4.6 Sentinel-1 and Sentinel-2 multimodal classification

A longer-term extension would combine Sentinel-1 SAR and Sentinel-2 optical information.

The two sensors provide complementary measurements.

Sentinel-1 represents microwave scattering and remains operational under cloud cover, while Sentinel-2 provides spectral information directly related to vegetation reflectance, soil properties, water, and built surfaces.

Multi-source deep learning studies have shown that Sentinel-1 and Sentinel-2 time series can provide complementary information for land-cover mapping [49].

For the Philippines, this direction could be particularly useful for the Forest–Grassland–Shrubland problem and for Mangrove discrimination.

However, such a multimodal experiment should be treated as an extension of the present SAR-only work rather than as a replacement for the current baseline.

7.4.7 Geographically aware training and evaluation

The Shrubland analysis indicates that class appearance can vary between geographic areas.

Future dataset design should therefore consider geography explicitly.

Possible approaches include:

- region-balanced training batches;
- leave-one-region-out validation;
- geographically disjoint test sets;
- domain-aware sampling;
- region-specific normalisation or adaptation;
- domain-generalisation methods.

The objective would not be to maximise accuracy on already represented locations, but to improve performance on areas whose feature distribution differs from the training sample.

This is consistent with current research on land-cover mapping under domain shift [48].

7.4.8 Improved treatment of rare classes

The negative result of the 10× sampler does not imply that rare-class learning should be abandoned.

Instead, future experiments should be more targeted.

Potential directions include:

- class-aware patch selection based on both class amount and spatial coherence;
- focal or hybrid losses;
- region-balanced sampling;
- hierarchical vegetation classification;
- contrastive feature learning;
- class prototypes;
- uncertainty-aware label weighting;
- targeted augmentation of validated rare-class examples.

These methods should be evaluated one at a time against the clean CC50 baseline.

The present sampler experiment demonstrates why several modifications should not be introduced simultaneously: if accuracy changes, the cause must remain identifiable.

7.4.9 Hierarchical land-cover classification

The confusion matrix suggests that some errors occur inside semantically related groups.

For example:

$$\text{Forest} \leftrightarrow \text{Grassland} \leftrightarrow \text{Shrubland} \tag{7.49}$$

represent one difficult vegetation group, while:

$$\text{Water} \leftrightarrow \text{Wetland} \tag{7.50}$$

represent another.

A hierarchical classifier could first separate broad surface types and then apply a more specialised classifier inside the difficult group.

For example:

$$\text{Vegetated} \rightarrow \{\text{Forest}, \text{Grassland}, \text{Shrubland}\}. \tag{7.51}$$

This strategy may be more appropriate when the between-group differences are strong but within-group differences are subtle.

7.4.10 Reference confidence and provenance

Another promising direction is to retain more information about how each reference label was generated.

Instead of storing only the final class ID, each pixel could also be associated with information such as:

- number of agreeing source products;
- identities of agreeing products;
- original versus rescued provenance;
- component size;
- agreement confidence.

This would allow the training loss to distinguish between very high-confidence reference pixels and pixels obtained through a weaker agreement condition.

Such provenance information was already useful during the Shrubland diagnostic and could become an explicit part of future training.

7.4.11 Full-tile mapping

The current thesis concentrates on the creation of the training/reference system and on controlled semantic-segmentation experiments over selected scenes.

A natural operational extension is systematic full-tile inference over complete Sentinel-2 grid cells and evaluation of the resulting maps.

This requires:

- tiled or sliding-window inference;

- consistent overlap handling;
- reconstruction of full-resolution probability maps;
- preservation of geospatial metadata;
- large-raster memory management;
- automated quality-control checks.

Such an extension would allow the trained classifier to move from an experimental validation framework toward operational large-area production.

7.4.12 Expansion of the geographic training sample

Increasing the geographic diversity of the training dataset is likely to be more useful than simply increasing the number of repeated samples from the same locations.

Future training data should preferentially target:

- regions poorly represented in the current dataset;
- rare classes;
- difficult class boundaries;
- areas where the current model shows high uncertainty;
- geographic environments whose Sentinel-1 feature distributions differ from the existing training scenes.

This strategy would directly address the cross-area variability observed in the present study.

7.4.13 Final perspective

The work presented in this thesis demonstrates that a Sentinel-1 SAR semantic-segmentation workflow can achieve strong internal validation results for several major land-cover classes in a complex tropical and archipelagic study area.

The final clean CC50 model achieved an Overall Accuracy of approximately 83.3% and strong IoU values for Water, Forest, and Cropland.

At the same time, the study demonstrates that a single global accuracy value does not fully describe model quality.

Rare classes exposed limitations that would have remained hidden if the analysis had stopped at Overall Accuracy.

The Shrubland investigation was particularly informative because it linked the deep learning results back to the reference-generation process and then forward again to training distribution and geographic feature variability.

The resulting workflow should therefore be understood not only as a trained classification model, but as a complete framework connecting reference-data generation, SAR feature preparation, semantic segmentation, accuracy assessment, and iterative data-quality analysis.

The principal contribution of the thesis is this integrated perspective.

A reliable land-cover classifier depends not only on the neural network, but also on what is considered ground reference, how classes are represented, where training examples are located, and whether the remote-sensing information available to the model is capable of separating the target categories consistently across space.

Within these constraints, the developed Philippine MOLCA–Sentinel-1– Swin-UNETR workflow provides a reproducible experimental baseline from which future large-area SAR land-cover mapping studies can be developed.

Appendix A

Complete MOLCA Harmonisation Tables

This appendix records the source-to-target class translations used to construct the Philippines MOLCA-derived reference. The purpose is to preserve, in one place, the exact thematic decisions underlying the agreement procedure described in Chapter **??**. The common target system contains one unresolved code (0, No-Data/AnA) and nine land-cover classes (1–9). The tables below are supporting reproducibility material; the scientific rationale for the individual products remains in Chapters 2 and 3.

A.1 Common target legend

ID	Target class	Interpretation
0	NoData / AnA	Unresolved, invalid, or deliberately unassigned reference pixel.
1	Built-up	Settlement and other built surfaces.
2	Cropland	Annual and perennial agricultural land.
3	Forest	Forest classes other than Mangrove when separately identified.
4	Grassland	Herbaceous grass-dominated land cover.
5	Shrubland	Brush/shrub-dominated land cover.
6	Wetland	Marshland, swamp, and herbaceous wetland.
7	Water	Inland/surface water and fishponds according to the project rules.
8	Bareland	Open, barren, or sparsely vegetated surfaces.
9	Mangrove	Philippine coastal mangrove class.

Table A.1: Final ten-code reference legend used throughout the Philippines workflow.

A.2 Original global and thematic products

Table A.2: Complete harmonisation of the six original MOLCA-style input products.

Dataset	Source value	Target ID	Target class	Role / interpretation
FROM-GLC 2017	0	0	NoData	No valid class.
FROM-GLC 2017	10	2	Cropland	General land-cover class.
FROM-GLC 2017	20	3	Forest	General land-cover class.
FROM-GLC 2017	30	4	Grassland	General land-cover class.
FROM-GLC 2017	40	5	Shrubland	General land-cover class.
FROM-GLC 2017	50	6	Wetland	General land-cover class.

Dataset	Source value	Target ID	Target class	Role / interpretation
FROM-GLC 2017	60	7	Water	General land-cover class.
FROM-GLC 2017	80	8	Bareland	General land-cover class.
FROM-GLC 2017	90	1	Built-up	General land-cover class.
GlobeLand30 2020	10	2	Cropland	Cultivated land.
GlobeLand30 2020	20	3	Forest	Forest.
GlobeLand30 2020	30	4	Grassland	Grassland.
GlobeLand30 2020	40	5	Shrubland	Shrubland.
GlobeLand30 2020	50	6	Wetland	Wetland.
GlobeLand30 2020	60	7	Water	Water bodies.
GlobeLand30 2020	70	0	NoData / AnA	Tundra; not retained in the Philippine target legend.
GlobeLand30 2020	80	1	Built-up	Artificial surfaces.
GlobeLand30 2020	90	8	Bareland	Bare land.
GlobeLand30 2020	100	0	NoData / AnA	Permanent snow/ice; not retained.
GlobeLand30 2020	255	0	NoData / AnA	Background/ocean/NoData.
WSF 2019	255	1	Built-up	Built-up evidence in the project raster.
WSF 2019	0	0	NoData / AnA	Non-built-up; no evidence for another class.
WSF 2019	1	0	NoData / AnA	Treated as NoData/noise according to the project clarification.

Dataset	Source value	Target ID	Target class	Role / interpretation
GHS-BUILT-S	1–100	1	Built-up	Any non-zero built-up-surface value used as Built-up evidence.
GHS-BUILT-S	0	0	NoData / AnA	No built-up evidence.
GHS-BUILT-S	NoData	0	NoData / AnA	Outside valid coverage.
GSW 2019	3	7	Water	Permanent-water evidence.
GSW 2019	2	6	Wetland	Seasonal water mapped to Wetland by the project harmonisation rule.
GSW 2019	1	0	NoData / AnA	Not-water; no water evidence.
GSW 2019	0	0	NoData / AnA	Source NoData.
JAXA FNF 2019	1	3	Forest	Forest evidence.
JAXA FNF 2019	2	0	NoData / AnA	Non-forest; not used as evidence for another land-cover class.
JAXA FNF 2019	3	7	Water	Water evidence.
JAXA FNF 2019	0	0	NoData / AnA	Source NoData.

A.3 NAMRIA 2020 harmonisation

The NAMRIA source required an explicit raster unique-value audit after polygon merging, rasterisation, and VRT preparation. The final project mapping therefore used the observed raster values rather than assuming that the published category order directly matched raster codes. Source value 3 appeared in the audit but was not assigned a target land-cover class in the final mapping.

Value	NAMRIA class	ID	Target	Note
1	Closed Forest	3	Forest	Forest evidence.
4	Open Forest	3	Forest	Forest evidence.
7	Mangrove Forest	9	Mangrove	Retained as a separate Philippine class.
10	Brush/Shrubs	5	Shrubland	Direct thematic correspondence.
13	Open/Barren	8	Bareland	Open/barren land.
14	Grassland	4	Grassland	Direct thematic correspondence.
15	Marshland/Swamp	6	Wetland	Wetland class.
16	Annual Crop	2	Cropland	Agricultural class.
17	Perennial Crop	2	Cropland	Agricultural class.
19	Fishpond	7	Water	Project decision.
20	Built-up	1	Built-up	Built surfaces.
21	Inland Water	7	Water	Water class.
3	Unassigned in final mapping	0	NoData / AnA	Present in the raster audit but not used as target-class evidence.

Table A.3: NAMRIA 2020 raster-value mapping used in the final harmonisation.

A.4 JAXA Southeast Asia harmonisation

Table A.4: JAXA 2023SEA v25.09 harmonisation used in the Philippine workflow.

Value	JAXA class	ID	Target	Note
0	No Data	0	NoData / AnA	No valid evidence.
1	Water Bodies	7	Water	Direct mapping.
2	Built-up	1	Built-up	Direct mapping.
3	Solar Panel	1	Built-up	Project harmonisation.
4	Cropland	2	Cropland	Direct mapping.
5	Single-crop Paddy Field	2	Cropland	Agricultural mapping.
6	Multi-crop Paddy Field	2	Cropland	Agricultural mapping.
7	Herbaceous Wetland	6	Wetland	Direct mapping.

Value	JAXA class	ID	Target	Note
8	Grassland	4	Grassland	Direct mapping.
9	Deciduous Broad-leaved Forest	3	Forest	Forest class.
10	Evergreen Broad-leaved Forest	3	Forest	Forest class.
11	Evergreen Needle-leaved Forest	3	Forest	Forest class.
12	Rubber Tree Plantation	3	Forest	Supervisor-approved project decision.
13	Oil Palm Tree Plantation	3	Forest	Supervisor-approved project decision.
14	Mangrove Forest	9	Mangrove	Retained separately.
15	Bare	8	Bareland	Direct mapping.

Appendix B

MOLCA Processing, Quality Control, and Validation Evidence

This appendix supplements Chapter ?? with compact implementation logic and the numerical quality-control evidence preserved in the project archive. It is intentionally more concrete than the methodological chapter: it records the order of operations, the staged reduction of unresolved pixels, and the status of the available validation workbooks.

B.1 Tile-level processing sequence

For each Sentinel-2 tile, the reference-generation pipeline followed the same high-level sequence:

1. identify the tile extent and prepare source-specific master rasters/VRTs;
2. clip each source product to the tile processing extent;
3. translate source classes into the common ten-code legend;
4. align categorical rasters to the common 10 m reference grid with nearest-neighbour resampling;
5. inspect unique values and spatial correspondence;
6. compute the strict FROM-GLC/GlobeLand30 base agreement;
7. fill only still-unresolved pixels with Built-up evidence;
8. fill only still-unresolved pixels with Water/Wetland evidence;

9. fill only still-unresolved pixels with Forest evidence;
10. introduce Mangrove where the accepted regional evidence supports it;
11. apply final NAMRIA/JAXA regional agreement to remaining unresolved pixels;
12. export the final reference raster and the stage-wise AnA/NaN report.

The essential conservative rule is that later evidence stages do not overwrite a valid class assigned by an earlier accepted stage. In simplified form:

```
M = base_agreement(FROM_GLC, GlobeLand30)
M = fill_unresolved(M, built_up_evidence)
M = fill_unresolved(M, water_wetland_evidence)
M = fill_unresolved(M, forest_evidence)
M = fill_unresolved(M, mangrove_evidence)
M = fill_unresolved(M, regional_agreement(NAMRIA, JAXA_SEA))
export(M, stage_report)
```

This pseudocode is a reproducibility summary of the implemented logic, not a replacement for the archived processing scripts.

B.2 Stage-wise AnA reduction

The archived CSV reports use the field names `NaN Pixels` and `NaN Percentage`. In the thesis, these values are interpreted as unresolved AnA pixels after each stage.

Table B.1: Archived staged AnA/NaN reduction for four representative Sentinel-2 tiles.

Tile	**Stage**	**Classified pixels**	**AnA pixels**	**AnA (%)**
51PTR	Base Agreement	33,226,538	87,333,862	72.44
51PTR	Built-up	58,619,028	61,941,372	51.38
51PTR	Water/Wetland	111,186,344	9,374,056	7.78
51PTR	Forest	112,343,817	8,216,583	6.82
51PTR	Mangrove	112,379,338	8,181,062	6.79
51PTR	Final MOLCA	113,986,830	6,573,570	5.45
50PRB	Base Agreement	40,684,489	79,875,911	66.25
50PRB	Built-up	65,282,695	55,277,705	45.85

Tile	Stage	Classified pixels	AnA pixels	AnA (%)
50PRB	Water/Wetland	102,979,043	17,581,357	14.58
50PRB	Forest	105,831,793	14,728,607	12.22
50PRB	Mangrove	105,864,296	14,696,104	12.19
50PRB	Final MOLCA	110,319,026	10,241,374	8.49
51PUR	Base Agreement	47,142,590	73,417,810	60.90
51PUR	Built-up	68,367,455	52,192,945	43.29
51PUR	Water/Wetland	111,580,220	8,980,180	7.45
51PUR	Forest	112,430,255	8,130,145	6.74
51PUR	Mangrove	112,708,153	7,852,247	6.51
51PUR	Final MOLCA	113,884,894	6,675,506	5.54
51PUS	Base Agreement	47,703,306	72,857,094	60.43
51PUS	Built-up	67,705,640	52,854,760	43.84
51PUS	Water/Wetland	112,156,832	8,403,568	6.97
51PUS	Forest	113,454,886	7,105,514	5.89
51PUS	Mangrove	113,996,491	6,563,909	5.44
51PUS	Final MOLCA	114,504,872	6,055,528	5.02

The four examples show the same general pattern: the strict base agreement leaves a large unresolved area, while thematic evidence and regional agreement progressively increase labelled coverage. The final unresolved fraction is 5.45% for 51PTR, 8.49% for 50PRB, 5.54% for 51PUR, and 5.02% for 51PUS.

B.3 Logic-validation workbooks

Three validation workbooks are preserved in the project archive. Their roles are not identical and should not be merged into a single accuracy statement.

Tile	Sampled points	VALID	INVALID	Interpretation
51NXG	800	791	9	98.875% of the recorded logic checks were marked VALID.
51PXP	800	793	7	99.125% of the recorded logic checks were marked VALID.
51PUS	400	–	–	The archived workbook contains the 400-point class-stratified sampling structure, but no completed VALID/INVALID field from which a final percentage can be computed.

Table B.2: Status of the archived MOLCA logic-validation workbooks.

These checks verify implementation logic at sampled locations. They are not an independent thematic-accuracy assessment of the complete Philippine reference map, and they are therefore reported separately from the deep-learning validation metrics.

Appendix C

Patch Selection and Sentinel-1 Feature Preparation Evidence

This appendix records the numerical and file-organisation evidence supporting Chapter **??**. It connects the country-scale MOLCA-derived reference to the final model-ready deep-learning dataset.

C.1 From selected patches to model-ready scenes

Stage	Count	Meaning
Operational Sentinel-2 tiles	60	Tiles retained for the main Philippine workflow.
Candidate grid cells	24,000	60 tiles × 20 × 20 candidate grid.
Selected reference patches	280	Geographically and thematically designed patch sample.
Raw Sentinel-1 feature exports	7,840	280 patches × 28 feature layers.
Excluded selected patches	3	Three 51QUC patches without the complete required 2019 VH feature set.
Final model-ready scene–reference associations	277	Samples retained for deep learning.
Training scenes	249	Fixed training split.
Validation scenes	28	Fixed internal validation split.
Training network patches	2,241	Nine overlapping 256 × 256 patches per training scene.
Validation network patches	252	Nine overlapping patches per validation scene.

Table C.1: Main sample counts from spatial selection to deep-learning input.

The 280 selected geographic patches remain part of the sampling history. The final learning dataset contains 277 associations because missing Sentinel-1 inputs were not synthetically filled.

C.2 Final GEE-ready patch representation

The final patch-selection table uses the following fields:

Field	Meaning
`tile`	Sentinel-2 MGRS tile identifier.
`patch_id`	Unique TILE–row–column identifier.
`xmin`	Western longitude of the patch bounding box.
`ymin`	Southern latitude of the patch bounding box.
`xmax`	Eastern longitude of the patch bounding box.
`ymax`	Northern latitude of the patch bounding box.

Table C.2: Fields used to reconstruct selected patch geometries in Google Earth Engine.

A representative patch identifier is `51PTR_R09_C15`. The same spatial identifier is used to associate the ground-reference raster, Sentinel-1 features, and JSON label metadata.

C.3 Sentinel-1 acquisition and feature configuration

Item	Final configuration
Collection	Copernicus Sentinel-1 GRD.
Observation year	2019.
Orbit pass	Descending.
Instrument mode	IW.
Polarisation	VH.
Temporal groups	Jan–Mar, Apr–Jun, Jul–Sep, Oct–Dec.
Features per group	SI, LEE, MAX, MIN, MAXMIN, MEAN, MEDIAN.
Channels per scene	$4 \times 7 = 28$.
Local reference size before DL resize	549×549 pixels.
Model scene size	512×512 pixels.
Network patch size	256×256 pixels.
Stride	128 pixels (50% overlap).

Table C.3: Final Sentinel-1 and spatial configuration carried into the model.

The scientific description uses neutral temporal-group names rather than temperate season labels. The four groups preserve the implemented calendar boundaries while avoiding climatologically inappropriate terminology for the Philippines.

C.4 Model-ready directory organisation

The final dataset is organised around three linked components:

```
Final_Output/
  Ground_Reference/   categorical reference rasters
  Labels/             JSON scene/class metadata
  S1/                 28 aligned S1 features per scene
```

The categorical ground-reference rasters are handled with nearest-neighbour logic, whereas Sentinel-1 feature layers remain continuous floating-point predictor data. The GEE exports are intermediate products; local reprojection, cropping, and alignment are required before the final 549×549 scene–reference pairing.

Appendix D

Deep-Learning Reproducibility Manifest

The final experiments were organised so that the clean CC50 baseline and the 10× sampler experiment are separable and reproducible. This appendix records the principal configuration values and the frozen-file evidence used during the audit.

D.1 Final clean-baseline configuration

Parameter	Value
Input channels	28 Sentinel-1 VH feature channels.
Output codes	10 (NoData ID 0 plus nine land-cover classes).
Model	2-D SwinUNETR implemented in MONAI.
Training scenes	249.
Validation scenes	28.
Batch size	1.
Maximum epochs	100.
Optimizer	Adam.
Initial learning rate	1×10^{-4}.
Weight decay	0.01.
Loss	Class-weighted categorical cross-entropy.
Ignored target	NoData (ID 0).
Monitoring metric	Validation mIoU.
Early-stopping patience	25 epochs.
Minimum improvement	0.001.
Best clean-baseline checkpoint	Epoch 53.
Clean-baseline stop epoch	78.

Table D.1: Principal frozen configuration of the final clean CC50 baseline.

D.2 Final CC50 class weights

ID	Class	Weight
0	NoData	0.0000
1	Built-up	0.6563
2	Cropland	0.2637
3	Forest	0.1965
4	Grassland	0.5624
5	Shrubland	3.5999
6	Wetland	0.8742
7	Water	0.3042
8	Bareland	0.7373
9	Mangrove Forest	1.8056

Table D.2: Class weights frozen for the final CC50 experiments.

The clean baseline and the sampler experiment use the same pixel-level loss weights. The sampler experiment changes training-patch exposure; it does not replace the weighted loss.

D.3 Frozen-file checksums

The project audit verified that the current split/statistics/weight files and training source matched the snapshots preserved with the clean baseline. The SHA-256 values recorded during that verification are:

Artifact	SHA-256
`tile_split.json`	`f6988d7bc2604306c37e089aeed3a7682fe2e316be93f173c7246d8ceb6e0e4c`
`class_weights.json`	`7dea0e7c8b69be0c5c39d554e297134722a269319c048ab511e946d820a0a313`
`tile_stats.json`	`c180719f4d5a2427c61d7eeb642b6c4ebac4faa6ee8d0ac3e94d377975cfb026`
`train.py`	`e1e569fe89596fabb7ee4d1d2c1c75d6d1591abe816cc04abb908dce2cd57d8b`

Table D.3: Checksums recorded for the frozen clean-baseline reproducibility package.

The audit also confirmed that `dataset.yaml`, `training.yaml`, and `legend.yaml` were identical to the corresponding clean-baseline snapshots.

D.4 Experiment-artifact organisation

The preserved results are separated into explicit experiment directories:

```
outputs/
  cc50_clean_baseline/
    config_snapshot/
    evaluation/
    training.log
    CC50_CLEAN_BASELINE_SUMMARY.txt

  cc50_weighted_sampler_10x/
    config_snapshot/
    evaluation/
    training.log
    SAMPLER_AUDIT.txt

  cc50_comparison/
```

```
    clean_vs_sampler10x_metrics.csv
    CC50_CLEAN_VS_SAMPLER10X_SUMMARY.txt

  cc50_pretraining_checks/
    rare-class / split / label audit tables
```

The clean and sampler evaluation directories preserve confusion matrices, per-class metrics, per-tile metrics, Shrubland validation summaries, and global evaluation summaries. Model checkpoints are stored separately on the lab/data storage because each checkpoint is approximately 302 MB.

Appendix E

Supplementary Evaluation Tables

This appendix consolidates the final numerical evaluation outputs so that the exact reference values remain available without searching across several figures in Chapter 6. These tables do not introduce a new experiment; they reproduce the frozen final results in compact form.

E.1 Clean baseline versus 10× sampler

Metric	CC50 clean	Sampler 10×	Difference
Best patch-level val. mIoU	0.369677	0.352474	-0.017203
Overall Accuracy	0.833297	0.803817	-0.029480
Macro precision	0.549740	0.505413	-0.044327
Macro recall	0.583169	0.551970	-0.031199
Macro F1	0.546032	0.512222	-0.033810
Full-scene mean IoU	0.433949	0.403811	-0.030138
Cohen's κ	0.770260	0.732253	-0.038007

Table E.1: Controlled comparison of the final clean CC50 baseline and 10× sampler experiment.

The sampler increased the expected probability of drawing a Shrubland-containing training patch from approximately 13.2% to 60.35%, but every reported global metric decreased. It is therefore retained as a controlled negative experiment rather than selected as the final model.

E.2 Per-class clean-baseline results

Table E.2: Per-class unique-pixel validation results for the final clean CC50 baseline.

Class	Support	Pred.	Prec.	Recall	F1	IoU
Built-up	173,461	187,193	0.5212	0.5624	0.5410	0.3708
Cropland	1,491,586	1,570,323	0.8424	0.8869	0.8641	0.7607
Forest	2,546,505	2,588,476	0.8879	0.9025	0.8951	0.8102
Grassland	431,429	275,000	0.6154	0.3923	0.4791	0.3150
Shrubland	3,716	0	0.0000	0.0000	0.0000	0.0000
Wetland	197,457	198,534	0.4852	0.4878	0.4865	0.3215
Water	1,128,402	1,158,187	0.9316	0.9561	0.9437	0.8934
Bareland	189,707	148,696	0.4960	0.3888	0.4359	0.2787
Mangrove Forest	11,963	47,817	0.1680	0.6717	0.2688	0.1553

NoData is excluded from the accuracy assessment. The table shows why Overall Accuracy alone is insufficient: Forest, Water, and Cropland are comparatively strong, whereas Shrubland is not recovered by the clean model and several minority classes remain substantially weaker.

E.3 Representative scene-level results

Scene	OA	Macro F1	mIoU
51QVA_14_06	0.9683	0.6198	0.5195
51QTU_01_10	0.8226	0.5995	0.4540
51QTV_09_10	0.8061	0.5596	0.4353
51QVV_12_03	0.9207	0.5125	0.4253
51PWM_09_03	0.9410	0.4411	0.3848
50PRC_11_05	0.7398	0.2564	0.1988
50PRC_20_06	0.4616	0.3859	0.2625

Table E.3: Representative full-scene validation results illustrating geographic variation in difficulty.

E.4 Shrubland evidence retained for interpretation

The final CC50 model-space training set contains 62,501 Shrubland pixels across 60 training scenes; the validation set contains 3,716 Shrubland pixels across five validation scenes. Among the 2,241 overlapping training patches, 296 contain at least one Shrubland pixel. In the validation set, 19 of 252 patches contain Shrubland.

Of the 62,501 training Shrubland pixels, 9,276 (14.84%) were inherited from the original MOLCA reference and 53,225 (85.16%) were introduced by the conservative CC50 rescue. This provenance is central to the interpretation in Chapter 6: the reference revision substantially increased available class-5 supervision, but the model still did not form a reliable Shrubland decision region.

Appendix F

Project Evidence and Reproducibility File Manifest

The following files and artefact groups form the evidence base used to reconstruct, audit, and interpret the workflow. This appendix documents the project provenance without embedding private credentials, large binary checkpoints, or redundant copies of source code in the thesis PDF.

Table F.1: Principal project evidence retained outside the main thesis narrative.

Artifact	**Purpose**
`6_datasets.xlsx`	Source-value and harmonisation evidence for the original MOLCA-style products.
`NAMRIAHarmonisation.xlsx`	National-product raster values and final class correspondence.
`JAXAHarmonisation.xlsx`	JAXA Southeast Asia value-to-target-class correspondence.
`51PTR_NaN_Reduction_Report.csv`	Stage-wise reference coverage report for 51PTR.
`50PRB_NaN_Reduction_Report.csv`	Stage-wise reference coverage report for 50PRB.
`51PUR_NaN_Reduction_Report.csv`	Stage-wise reference coverage report for 51PUR.
`51PUS_NaN_Reduction_Report.csv`	Stage-wise reference coverage report for 51PUS.
`51NXG_Validation.xlsx`	800-point logic-validation table.
`51PXP_Validation.xlsx`	800-point logic-validation table.
`validation_51PUS_400.xlsx`	400-point class-stratified validation structure for pilot checking.

Artifact	Purpose
Final_GEE_Table_280_Patches.csv	Final selected-patch extents used for Sentinel-1 extraction.
60_Tiles_name.csv	Operational Sentinel-2 tile set.
Checklist_Tiles_7840_Patches.xlsx	Tracking of the planned 280 × 28 Sentinel-1 feature exports.
tile_split.json	Frozen 249/28 scene split used in the final experiments.
tile_stats.json	Scene-level class statistics used by the split/weighting workflow.
class_weights.json	Final CC50 inverse-square-root training weights.
training.yaml	Final clean-baseline training configuration.
dataset.yaml	Final Philippines CC50 dataset configuration.
legend.yaml	Ten-code Philippines reference/model legend.
evaluation_summary.json	Aggregate full-scene evaluation metrics.
per_class_metrics.csv	Final class-wise precision, recall, F1, IoU, support, and prediction counts.
per_tile_metrics.csv	Scene-level validation metrics.
shrubland_validation_tiles.csv	Shrubland-specific validation-scene evidence.
confusion_matrix_counts.csv	Global confusion-matrix counts.
confusion_matrix_row_normalized.csv	Row-normalised confusion matrix used for class-wise interpretation.
training.log	Epoch-level training record including monitored validation metrics and stopping behaviour.

Private notification tokens/credentials used during operational automation are not part of the scientific method and are intentionally excluded from the thesis and any public code package. Large trained-model checkpoints are likewise referenced by the reproducibility package but retained on project storage rather than embedded in the PDF.

The appendices therefore serve two purposes: they preserve the exact numerical and thematic decisions required to understand the reported results, and they identify the external artefacts required for full computational reproduction without turning the thesis into a source-code archive.

Bibliography

[1] Antonio Di Gregorio. *Land Cover Classification System: Classification Concepts and User Manual*. 2nd ed. Environment and Natural Resources Service Series 8. Rome, Italy: Food and Agriculture Organization of the United Nations, 2005.

[2] European Space Agency. *High Resolution Land Cover*. ESA Climate Change Initiative. 2026. URL: https://climate.esa.int/en/projects/high-resolution-land-cover/ (visited on 08/19/2026).

[3] Gorica Bratic, Daniele Oxoli, and Maria Antonia Brovelli. "Map of Land Cover Agreement: Ensambling Existing Datasets for Large-Scale Training Data Provision". In: *Remote Sensing* 15.15 (2023), p. 3774. DOI: 10.3390/rs15153774.

[4] Segun Ajibola and Pedro Cabral. "A Systematic Literature Review and Bibliometric Analysis of Semantic Segmentation Models in Land Cover Mapping". In: *Remote Sensing* 16.12 (2024), p. 2222. DOI: 10.3390/rs16122222.

[5] Ramon Torres et al. "GMES Sentinel-1 Mission". In: *Remote Sensing of Environment* 120 (2012), pp. 9–24. DOI: 10.1016/j.rse.2011.05.028.

[6] Alberto Moreira et al. "A Tutorial on Synthetic Aperture Radar". In: *IEEE Geoscience and Remote Sensing Magazine* 1.1 (2013), pp. 6–43. DOI: 10.1109/MGRS.2013.2248301.

[7] Maïlys Lopes et al. "Improving the Accuracy of Land Cover Classification in Cloud Persistent Areas Using Optical and Radar Satellite Image Time Series". In: *Methods in Ecology and Evolution* 11.4 (2020), pp. 532–541. DOI: 10.1111/2041-210X.13359.

[8] Lin Zhao and Shunping Ji. "CNN, RNN, or ViT? An Evaluation of Different Deep Learning Architectures for Spatio-Temporal Representation of Sentinel Time Series". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16 (2023), pp. 44–56. DOI: 10.1109/JSTARS.2022.3219816.

[9] Luigi Russo et al. “A Deep Learning Architecture for Land Cover Mapping Using Spatio-Temporal Sentinel-1 Features”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 18 (2025), pp. 10562–10581. DOI: 10.1109/JSTARS.2025.3557687.

[10] Nico R. Almarines et al. “Influence of Image Compositing and Multisource Data Fusion on Multitemporal Land Cover Mapping of Two Philippine Watersheds”. In: *Remote Sensing* 16.12 (2024), p. 2167. DOI: 10.3390/rs16122167.

[11] Philippine Atmospheric, Geophysical and Astronomical Services Administration. *Climate of the Philippines*. DOST-PAGASA. 2026. URL: https://www.pagasa.dost.gov.ph/information/climate-philippines (visited on 08/19/2026).

[12] Jonathan Long, Evan Shelhamer, and Trevor Darrell. “Fully Convolutional Networks for Semantic Segmentation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Vol. 9351. Lecture Notes in Computer Science. Springer, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[14] Ze Liu et al. “Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 10012–10022. DOI: 10.1109/ICCV48922.2021.00986.

[15] Ali Hatamizadeh et al. “Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images”. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer, 2022, pp. 272–284. DOI: 10.1007/978-3-031-08999-2_22.

[16] M. Jorge Cardoso, Wenqi Li, Richard Brown, et al. “MONAI: An Open-Source Framework for Deep Learning in Healthcare”. In: *arXiv preprint arXiv:2211.02701* (2022). DOI: 10.48550/arXiv.2211.02701.

[17] J. Solórzano et al. “Land Use Land Cover Classification with U-Net: Advantages of Combining Sentinel-1 and Sentinel-2 Imagery”. In: *Remote Sensing* 13.18 (2021), p. 3600. DOI: 10.3390/rs13183600.

[18] GADM. *GADM Database of Global Administrative Areas, Version 4.1*. Geospatial boundary dataset. 2022. URL: https://gadm.org/.

[19] European Space Agency. *Sentinel-2 Products Specification Document*. Copernicus Sentinel technical documentation. 2025. URL: https://sentinels.copernicus.eu/.

[20] Copernicus SentiWiki. *Sentinel-2 Products*. Copernicus Sentinel documentation. 2026. URL: https://sentiwiki.copernicus.eu/web/s2-products.

[21] Peng Gong et al. "Stable Classification with Limited Sample: Transferring a 30-m Resolution Sample Set Collected in 2015 to Mapping 10-m Resolution Global Land Cover in 2017". In: *Science Bulletin* 64.6 (2019), pp. 370–373. DOI: 10.1016/j.scib.2019.03.002.

[22] Jun Chen et al. "Global Land Cover Mapping at 30 m Resolution: A POK-Based Operational Approach". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 103 (2015), pp. 7–27. DOI: 10.1016/j.isprsjprs.2014.09.002.

[23] Martino Pesaresi and Panagiotis Politis. *GHS-BUILT-S R2023A: GHS Built-Up Surface Grid, Derived from Sentinel-2 Composite and Landsat, Multitemporal (1975–2030)*. European Commission, Joint Research Centre, 2023. DOI: 10.2905/9F06F36F-4B11-47EC-ABB0-4F8B7B1D72EA. URL: https://human-settlement.emergency.copernicus.eu/ghs_buS2023.php.

[24] German Aerospace Center (DLR). *World Settlement Footprint (WSF) 2019*. DLR Earth Observation Center dataset. 2021. DOI: 10.15489/twg5xsnquw84. URL: https://geoservice.dlr.de/web/datasets/wsf_2019.

[25] Jean-François Pekel et al. "High-resolution mapping of global surface water and its long-term changes". In: *Nature* 540.7633 (2016), pp. 418–422. DOI: 10.1038/nature20584.

[26] European Commission Joint Research Centre and Google. *JRC Yearly Water Classification History, v1.4*. Google Earth Engine Data Catalog. Accessed 30 September 2026. URL: https://developers.google.com/earth-engine/datasets/catalog/JRC_GSW1_4_YearlyHistory.

[27] Masanobu Shimada et al. "New Global Forest/Non-Forest Maps from ALOS PALSAR Data (2007–2010)". In: *Remote Sensing of Environment* 155 (2014), pp. 13–31. DOI: 10.1016/j.rse.2014.04.014.

[28] Japan Aerospace Exploration Agency. *Global 25 m Resolution PALSAR-2 Forest/Non-Forest Map*. JAXA Earth Observation Research Center dataset. 2019. URL: https://www.eorc.jaxa.jp/ALOS/en/dataset/fnf_e.htm.

[29] Philippine Statistics Authority. *The Philippines' Open and Closed Forest Cover Grew by 2.9 Percent in 2020*. Official statistical release. 2024. URL: https://psa.gov.ph/content/philippines-open-and-closed-forest-cover-grew-29-percent-2020.

[30] National Mapping and Resource Information Authority. *NAMRIA Infomapper 2025*. Official NAMRIA publication. 2025. URL: https://www.namria.gov.ph/jdownloads/Info_Mapper/NAMRIA_Infomapper_2025.pdf.

[31] Japan Aerospace Exploration Agency. *High-Resolution Land-Use and Land-Cover Map of Southeast Asia in 2023 (2023SEA_v25.09)*. JAXA Earth Observation Research Center dataset documentation. Released in September 2025, Version 25.09. 2025. URL: https://www.eorc.jaxa.jp/ALOS/en/dataset/lulc/lulc_v2509_e.htm.

[32] Copernicus Data Space Ecosystem. *Sentinel-1 Documentation*. Copernicus Data Space Ecosystem documentation. 2026. URL: https://documentation.dataspace.copernicus.eu/Data/Sentinel1.html.

[33] European Space Agency. *Introducing the Sentinel-1 Mission*. ESA mission documentation. 2026. URL: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1.

[34] Jong-Sen Lee. "Speckle analysis and smoothing of synthetic aperture radar images". In: *Computer Graphics and Image Processing* 17.1 (1981), pp. 24–32. DOI: 10.1016/S0146-664X(81)80005-6.

[35] Hu Cao et al. "Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation". In: *arXiv preprint arXiv:2105.05537* (2021). arXiv: 2105.05537 [eess.IV].

[36] Haibo He and Edwardo A. Garcia. "Learning from Imbalanced Data". In: *IEEE Transactions on Knowledge and Data Engineering* 21.9 (2009), pp. 1263–1284. DOI: 10.1109/TKDE.2008.239.

[37] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *arXiv preprint arXiv:1412.6980* (2014). DOI: 10.48550/arXiv.1412.6980.

[38] Ilya Loshchilov and Frank Hutter. "SGDR: Stochastic Gradient Descent with Warm Restarts". In: *arXiv preprint arXiv:1608.03983* (2016). DOI: 10.48550/arXiv.1608.03983.

[39] Russell G. Congalton. "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data". In: *Remote Sensing of Environment* 37.1 (1991), pp. 35–46. DOI: 10.1016/0034-4257(91)90048-B.

[40] Pontus Olofsson et al. “Good Practices for Estimating Area and Assessing Accuracy of Land Change”. In: *Remote Sensing of Environment* 148 (2014), pp. 42–57. DOI: 10.1016/j.rse.2014.02.015.

[41] Jacob Cohen. “A Coefficient of Agreement for Nominal Scales”. In: *Educational and Psychological Measurement* 20.1 (1960), pp. 37–46. DOI: 10.1177/001316446002000104.

[42] B. Hosseiny et al. “WetNet: A Spatial-Temporal Ensemble Deep Learning Model for Wetland Classification Using Sentinel-1 and Sentinel-2”. In: *IEEE Transactions on Geoscience and Remote Sensing* (2021). DOI: 10.1109/TGRS.2021.3113856.

[43] Zhen Shen et al. “Evaluating Feature Selection Methods and Machine Learning Algorithms for Mapping Mangrove Forests Using Optical and Synthetic Aperture Radar Data”. In: *Remote Sensing* 15 (2023), p. 5621. DOI: 10.3390/rs15235621.

[44] Johannes N. Hansen, Edward T. A. Mitchard, and Stuart King. “Assessing Forest/Non-Forest Separability Using Sentinel-1 C-Band Synthetic Aperture Radar”. In: *Remote Sensing* 12.11 (2020), p. 1899. DOI: 10.3390/rs12111899.

[45] Pengdi Chen et al. “Class Imbalance in the Automatic Interpretation of Remote Sensing Images: A Review”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 18 (2025), pp. 9483–9508. DOI: 10.1109/JSTARS.2025.3555567.

[46] Sarah Farhadpour, Timothy A. Warner, and Aaron E. Maxwell. “Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices”. In: *Remote Sensing* 16 (2024), p. 533. DOI: 10.3390/rs16030533.

[47] Junjue Wang et al. “LoveDA: A Remote Sensing Land-Cover Dataset for Domain Adaptive Semantic Segmentation”. In: *arXiv preprint arXiv:2110.08733* (2021). DOI: 10.48550/arXiv.2110.08733.

[48] Chao Ji and Hong Tang. “Towards Reliable Land Cover Mapping under Domain Shift: An Overview and Comprehensive Comparative Study on Uncertainty Estimation”. In: *Earth-Science Reviews* (2025). DOI: 10.1016/j.earscirev.2025.105070.

[49] Dino Ienco et al. “Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for Land Cover Mapping via a Multi-Source Deep Learning Architecture”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 158 (2019), pp. 11–22. DOI: 10.1016/j.isprsjprs.2019.09.016.