



**POLITECNICO
MILANO 1863**

SCHOOL OF CIVIL
ENVIRONMENTAL
AND LAND MANAGEMENT
ENGINEERING

High-Resolution Land Cover Mapping via Consensus Fusion and GeoAI: A Case Study in the Philippines

TESI DI LAUREA MAGISTRALE IN
GEOINFORMATICS ENGINEERING - INGEGNERIA GEOINFORMATICA

Author: **Yongbin Bao**

Student ID: 10945066

Advisor: Prof. Vasil Yordanov

Co-advisors: Prof. Maria Antonia Brovelli, Eng. Qiongjie Xu

Academic Year: 2025-26

Abstract

High-resolution land cover (HRLC) maps are essential for diverse fields, but obtaining high-quality training data remains a major challenge. To address this, this study proposes an automated framework that integrates existing HRLCs and GeoAI to generate high-accuracy Land Cover (LC) maps in the Philippines. By fusing existing HRLCs, the Map of Land Cover Agreement (MOLCA) and the Map of Land Cover Partial Agreement (MOLCOPA) were generated. MOLCA serves as a benchmark LC map, and MOLCOPA provides pixel-wise class probabilities. Different strategies effectively integrating GeoAI approaches with MOLCA and MOLCOPA were implemented. These approaches utilized Google Satellite Embeddings (GSE), 64-dimensional feature vectors extracted from the AlphaEarth Foundation Model. Both unsupervised approach (K-Means clustering with auto-labeling) and semi-supervised classifiers (Random Forest, Support Vector Machine, and Multi-Layer Perceptron) were evaluated. The semi-supervised approach, specifically with Multi-Layer Perceptron (MLP), which extracts training samples from MOLCA and leverages the MOLCOPA priors, emerged as the most robust model. It achieved an Overall Accuracy (OA) of 80.86% and a Kappa Coefficient of 0.78 on an independent validation set, demonstrating the significant potential of the proposed framework for land cover mapping.

Keywords: Land Cover, High-resolution, Earth Observation, Benchmark, GeoAI, Google Satellite Embedding

Abstract in lingua italiana

Le mappe di copertura del suolo (Land Cover - LC) ad alta risoluzione (HRLC) sono essenziali per svariati settori, tuttavia l'acquisizione di dati di addestramento di alta qualità rimane una sfida significativa. Per far fronte a tale problema, il presente studio propone un framework automatizzato che integra le HRLC esistenti e la GeoAI per generare mappe di copertura del suolo (LC) ad elevata accuratezza nelle Filippine. Tramite la fusione delle HRLC esistenti, sono state generate la Map of Land Cover Agreement (MOLCA) e la Map of Land Cover Partial Agreement (MOLCOPA). La MOLCA funge da mappa LC di riferimento (benchmark), mentre la MOLCOPA fornisce le probabilità di classe a livello di pixel. Sono state implementate diverse strategie che integrano efficacemente approcci di GeoAI con MOLCA e MOLCOPA. Tali approcci hanno utilizzato i Google Satellite Embeddings (GSE), vettori di feature a 64 dimensioni estratti dal Foundation Model AlphaEarth. Sono stati valutati sia un approccio non supervisionato (clustering K-Means con auto-etichettatura) sia classificatori semi-supervisionati (Random Forest, Support Vector Machine e Multi-Layer Perceptron). L'approccio semi-supervisionato, nello specifico con il Multi-Layer Perceptron (MLP), che estrae i campioni di addestramento dalla MOLCA e sfrutta le informazioni a priori (priors) della MOLCOPA, è risultato essere il modello più robusto. Ha ottenuto un'Accuratezza Complessiva (OA) dell'80,86% e un Coefficiente Kappa di 0,78 su un set di validazione indipendente, dimostrando il significativo potenziale del framework proposto per la mappatura della copertura del suolo.

Parole chiave: Copertura del suolo, Alta risoluzione, Osservazione della Terra, Benchmark, GeoAI, Google Satellite Embedding

Contents

Abstract	i
Abstract in lingua italiana	iii
Contents	v
1 Introduction	1
1.1 Objective	2
1.2 Thesis Structure	3
2 Data Preparation	5
2.1 Study Area	5
2.2 HRLC Datasets	6
2.2.1 Multi-class HRLC	6
2.2.2 Binary HRLC Datasets	9
2.3 Google Satellite Embedding	12
2.4 External Validation Datasets	13
2.5 Data Acquisition	13
2.6 Legend Harmonization	16
2.7 Critical Analysis	19
2.7.1 Distribution of External Validation Datasets	20
2.7.2 Analysis on Multi-class HRLCs	21
3 Methodology	27
3.1 Overview	27
3.2 Introduction to Techniques	29
3.2.1 Tiling System	29
3.2.2 Machine Learning Algorithm	30
3.2.3 Other Techniques	31

3.3	Unanimous Consensus: MOLCA	31
3.3.1	Overall Workflow	31
3.3.2	Target Defining and Data Preparation	33
3.3.3	Unanimous Consensus Rules	35
3.3.4	Implementation	36
3.3.5	Limitations	37
3.4	Probabilistic Consensus: MOLCOPA	38
3.4.1	Overall Workflow	38
3.4.2	Target Defining and Data Preparation	40
3.4.3	Probabilistic Consensus Rules	40
3.4.4	Derivation of Land Cover Maps	41
3.4.5	Implementation	43
3.5	GeoAI-based Land Cover Mapping	45
3.5.1	Unsupervised Approach	45
3.5.2	Semi-supervised Approach	48
3.6	Evaluation	58
3.6.1	MOLCA and MOLCOPA	58
3.6.2	GeoAI-based Land Cover Maps	59
4	Results	63
4.1	MOLCA	63
4.1.1	Maps	63
4.1.2	Statistical Analysis	65
4.1.3	Accuracy Assessment	67
4.2	MOLCOPA	69
4.2.1	Maps	69
4.2.2	Summary	74
4.2.3	Statistical Analysis	74
4.2.4	Accuracy Assessment	77
4.3	GeoAI-based Land Cover Mapping	85
4.3.1	Unsupervised Approach	85
4.3.2	Semi-supervised Approach	91
4.3.3	Comprehensive Comparison	101
5	Conclusion	109
	Bibliography	111

A Appendix A	119
List of Figures	131
List of Tables	135
List of Symbols	137
Acknowledgements	139

1 | Introduction

Land cover is the observed (bio)physical cover on the Earth's surface [19]. Land Cover (LC) maps have various applications in different fields, such as hydrological modeling [43], climate change [20], and disease spread [22], which highlights the importance of LC maps.

Driven by advancements in computational power and the open access to Landsat and Sentinel data, a variety of High-resolution Land Cover (HRLC) products have emerged over the last decade (e.g., Dynamic World [6], GlobeLand30 [10], GLC_FCS30 [53], FROM-GLC30 [23], FROM-GLC10 [24], European Space Agency (ESA) WorldCover [49, 50], and ESRI LULC [30]). Seven mentioned HRLC products were all developed based on Machine Learning [10, 23, 24, 49, 50, 53] or Deep Learning [6, 30] techniques, which require a high volume of training data [47]. The collection of high-quality training data remains a major challenge in HRLC mapping.

Training data are commonly derived from manual visual interpretation, multi-resolution LC products, or a combination of both [3]. However, both approaches have significant drawbacks. Visual interpretation is a highly time-consuming task, while reusing existing LC data strongly depends on the reliability of the source maps and possibly introduces new errors due to legend harmonization issues.

To address this, G. Bratic et al. [3] proposed Map of Land Cover Agreement (MOLCA), a method that integrates existing HRLC maps to generate a benchmark land cover map based on unanimous consensus, from which training sample points can be extracted. However, MOLCA has certain limitations. It has a relatively small spatial coverage.

Considering the limitations of MOLCA, Map of Land Cover Partial Agreement (MOLCOPA) is proposed in this study. MOLCOPA introduces a probabilistic consensus based strategy that quantifies the pixel-wise probability for each class. A source-count map indicating the number of valid source multi-class maps per pixel is also produced with MOLCOPA, which can provide a confidence level for the MOLCOPA results. MOLCOPA and the count map can be utilized with various strategies to generate customized land cover maps with different confidence levels and coverages.

Nevertheless, although MOLCOPA significantly improves spatial coverage, achieving a 100% complete land cover map requires GeoAI-based approaches to obtain a complete LC map.

Traditional EO machine learning models heavily rely on optical spectral bands and vegetation indices. Meanwhile, embedding datasets extracted from foundation models, such as Google Satellite Embedding (GSE) [7], TESSERA [21], and PRESTO [46], are emerging as a transformative tool in the field of Earth Observation (EO). GSE [7] stands out due to its excellent spatial and temporal coverage. They are feature vectors that summarize relationships across multi-source, multi-modal observations in a less directly interpretable, but more powerful way [25]. This eliminates the hassle of manual feature selection in land cover mapping. In this study, various GeoAI-based approaches are employed, including unsupervised and semi-supervised learning approaches. These approaches effectively combine the embeddings with the previously generated MOLCA, MOLCOPA and the count map to produce a complete land cover map.

1.1. Objective

The core objective: Develop an automated and robust framework that leverages multiple existing HRLC datasets and GeoAI approaches to generate a complete and high-accuracy land cover map.

The detailed objectives are as follows.

- (1) To select and harmonize available HRLC products, and perform fundamental analyses on them.
- (2) To generate a benchmark land cover map (MOLCA) utilizing the unanimous consensus rule and quantify its limitations regarding spatial coverage.
- (3) To design and implement the probabilistic consensus approach fusing multiple existing HRLCs to provide pixel-wise class probabilities (MOLCOPA). And to explore various strategies to derive LC maps, discovering the trade-off between accuracy and spatial coverage.
- (4) To integrate GSE with MOLCA and MOLCOPA, generating a complete LC maps, followed by an accuracy assessment.

1.2. Thesis Structure

The remainder of this thesis is structured as follows.

Chapter 2 (Data Preparation) introduces the study area, the used HRLC datasets, the Google Satellite Embeddings, and an external validation dataset. It also details the legend harmonization rules and presents some analyses of the input datasets.

Chapter 3 (Methodology) details the theoretical and algorithmic implementations of the research, including MOLCA, MOLCOPA, the designed Geo-AI based LC mapping pipelines and the evaluation strategies.

Chapter 4 (Results) presents the outcomes of this study.

Chapter 5 (Conclusion) highlights the key findings of the research.

2 | Data Preparation

2.1. Study Area

As shown in Figure 2.1, the study area covers the terrestrial territory of the Philippines, which is an archipelagic country located in Southeast Asia on the western rim of the Pacific Ocean. It comprises approximately 7,641 islands with a total land area of about 300,000 km², of which 298,170 km² is land and 1,830 km² is water [8].

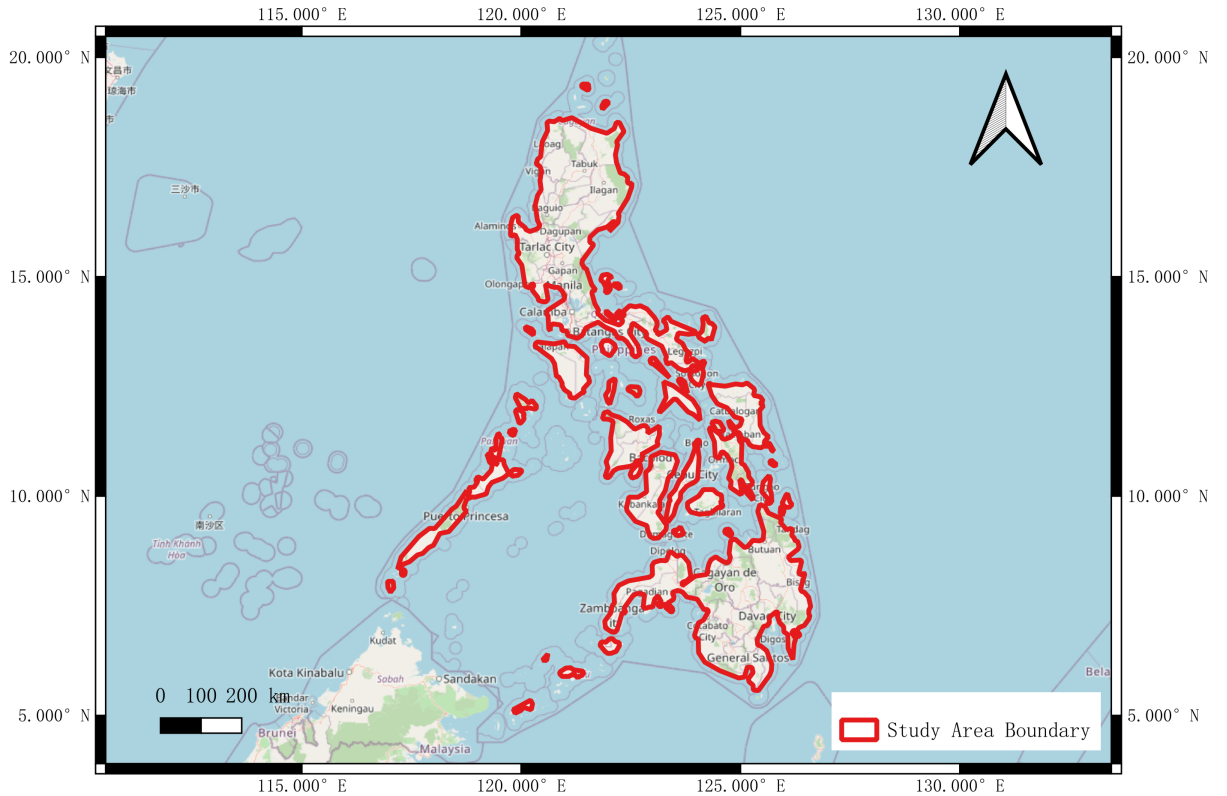


Figure 2.1: Study Area

2.2. HRLC Datasets

To generate maps for different years in this study, several High-Resolution Land Cover (HRLC) datasets were used. HRLC refers to those datasets that have a spatial resolution of at least 30 meters. Global HRLCs and regional HRLCs that cover the study area can be considered.

HRLC can be divided into two types: multi-class HRLC and binary HRLC. Multi-class HRLC refers to an HRLC data set that has a general purpose and classifies the land into multiple classes. Binary HRLC refers to an HRLC dataset that focuses on a specific field such as forest mapping and built-up mapping; typically this type has a binary value for each pixel on the map.

2.2.1. Multi-class HRLC

In this part, the multi-class HRLC datasets used in this study (as shown in Table 2.1) are introduced.

Dataset Name	Sensor	Algorithm	Temporal Resolution	Spatial Resolution
Dynamic World [6]	Sentinel-2	Deep Learning (FCNN)	from 2015-06-27 to present	10 m
Esri LULC [30]	Sentinel-2	Deep Learning (UNet)	Annual (2017–2023)	10 m
FROM-GLC10 [24]	Landsat 8, Sentinel-2	Random Forest	2017	10 m
FROM-GLC30 [23]	Landsat TM/ETM+	MLC/ RF / SVM/ J4.8	2010, 2015, 2017	30 m
GlobeLand30 [10]	Landsat TM/ETM+, HJ-1	POK	2000, 2010, 2019, 2025	30 m
GLC_FCS30D [53]	Landsat	Random Forest	Every 5 years (1985–1995) & Annual (2000–2022)	30 m
WorldCover [49, 50]	Sentinel-1 & 2	Gradient Boosting Decision Trees (CatBoost)	2020, 2021	10 m

Table 2.1: Multi-class HRLC Datasets

Dynamic World (DW) [6] developed by Google and other partners is the first near real-

time product with 10-meter resolution, which is a significant advancement for global land cover mapping. DW uses Google Earth Engine to process Sentinel-2 Level-1C Top-of-Atmosphere (TOA) images as L2A products are available from 2017 onwards but L1C has been generated since the beginning of Sentinel-2 project in 2015. Many older global land cover datasets use yearly image composites to reduce clouds. DW works with a strategy which enables it to create near real-time (NRT) land use land cover maps. It uses a deep fully convolutional neural network (FCNN) [44] to make land cover predictions each time new satellite images arrive. The applied FCNN is 100x smaller than other architectures used for ground-level camera imagery (e.g., U-net [41] and DeepLab v3 [11]). The model was trained with a semi-supervised method. The stratified training dataset included about 24,000 image tiles. Experts and non-experts labeled these tiles across 14 biomes, so that the dataset can represent the world well. DW also has an important feature: it does not only give one land cover label for each pixel. Instead, it outputs a probability for each of nine land cover classes. This helps users see how confident the model's prediction is and users can also create a customized land cover map product with the probability maps. This "soft classification" can describe transition areas and mixed pixels better than a single fixed class. In terms of the accuracy, DW was tested using an independent dataset made by expert consensus. It reached an overall accuracy of 73.8% [6].

The Esri LULC [30] developed by Impact Observatory, Esri and Microsoft provides annual land cover maps at 10-meter resolution from 2017–2023. This product was derived from Sentinel-2 L2A surface reflectance corrected imagery and produced with deep learning models U-Net. The U-Net model was trained using a novel dataset containing over 5 billion human-labeled pixels [30]. In terms of validation, a set of tiles labeled by experts was considered as a "gold standard" to do the validation. The results indicate that this product can achieve an overall agreement of 85% across all classes [30].

FROM-GLC30 [23] developed by Tsinghua University is the first global land cover product at 30-meter resolution derived from Landsat TM/ETM+ imagery. Before this work, global mapping was largely restricted to coarse-resolution sensors. They processed thousands of Landsat scenes using different supervised classifiers and compared several algorithms, including Maximum Likelihood, J4.8 decision trees, Random Forest, and Support Vector Machines (SVM). The evaluation indicated that SVM and Random Forest performed best for high-dimensional global mapping [23]. FROM-GLC30 adopts a hierarchical legend with 10 Level-1 classes. For the accuracy assessment, 38,664 test samples were manually collected by four image interpreters who did not participate in the collection of training samples. The result indicated that FROM-GLC30 can achieve the best overall accuracy of 64.9% with SVM classifier [23].

FROM-GLC10 [24] developed by Tsinghua University is a global 10 m land cover map for 2017. It was generated by adopting a novel strategy called "sample transfer." Instead of collecting a new sample set, they applied the samples collected from Landsat 8 images at 30-meter resolution for 2015 to train a Random Forest classifier with Sentinel-2 images at 10-meter resolution for 2017. It means that it was a cross-sensor, cross-resolution, cross-year sample transfer. The robustness of the classifier was also verified by the authors, less than 1% decrease of overall accuracy of land cover map could be observed when 20% of training samples were in error or only 40% of training samples were used. In terms of accuracy, FROM-GLC10 achieved 72.76% which is comparable to 2017 30-m global land cover map FROM-GLC30 but provides more spatial details [24]. The feasibility of sample transfer in global land cover mapping was proven [24].

GlobeLand30 (GL30) [10] was developed by the National Geomatics Center of China (NGCC), which utilized a classification system of 10 classes of land cover. A novel strategy called Pixel-Object-Knowledge (POK) was applied to generate the land cover map. POK is a hybrid methodology that integrates pixel-based, object-based and knowledge-based approaches, with the aim of improving reliability [10]. In addition, land cover mapping was decomposed into a series of per-class extraction with a certain order that starts from water bodies. For every class, hybrid pixel- and object-based classification was performed first, and the pixel-based method, the object-based method varies from class to class. After that, knowledge-based verification would be done to further improve the product quality. "Knowledge-based verification" refers to an interactive verification procedure, in which experts employ nature-based, culture-based, and temporal consistency rules to identify and correct classification errors, thereby ensuring the logical consistency and accuracy of the final land cover product [10]. For the accuracy assessment, a total of 159,874 pixel samples were collected across the world. The 2010 baseline product achieved an overall accuracy of 80.33%, demonstrating the efficacy of the POK strategy [10].

GLC_FCS30D [53] developed by Chinese Academy of Sciences (CAS) is the first global fine land cover dynamic product at a 30-meter resolution that adopts Continuous Change Detection (CCD) [55] which is a classical change-detection method based on dense time-series observations. A distinguishing feature of GLC_FCS30D is its Fine Classification System (FCS), which includes 35 detailed classes. A key advantage of GLC_FCS30D is that it provides land cover products from 1985-2022. Before 2000, the update cycle was every 5 years; after 2000, it is updated annually. Firstly, temporally stable areas were identified by exploiting CCD model and full time series of Landsat observations. The change-detection mask derived from CCD algorithm and GLC_FCS30-2020 prior dataset was utilized to generate spatiotemporally stable training samples. Random Forest was

selected as the classifier to be trained to update the changed areas due to its significant advantages in accommodating high-dimensional training features and excellent accuracy. In addition, a temporal-consistency optimization was performed to ensure the consistency of land cover maps. In terms of accuracy assessment, the product was validated with 84526 globally distributed samples from 2020; it achieved an overall accuracy of 80.88% for the basic classification system (10 major land cover classes) [53].

WorldCover [49, 50] developed by European Space Agency (ESA) provides global land-cover maps at 10 m resolution for 2020 (v100) and 2021 (v200). The product delivers a discrete land-cover map with 11 classes aligned with Food and Agriculture Organization of the United Nations (FAO) Land Cover Classification System (LCCS). Additional quality layers which contain per pixel quality indicators showing the quality of the Earth Observation (EO) input data are provided to users. Unlike datasets relying solely on optical imagery, WorldCover leverages the complementary strengths of both Sentinel-2 multi-spectral data and Sentinel-1 C-band Synthetic Aperture Radar (SAR). The classification workflow processes over 130 spectral-temporal features using the CatBoost gradient boosting algorithm [15]. In terms of accuracy assessment, the WorldCover product was validated using an independent validation dataset. The validation results showed that the overall accuracy of the WorldCover product is 74.4% for 2020 [49, 50].

2.2.2. Binary HRLC Datasets

Besides the above-mentioned Multi-class HRLC datasets, some binary HRLC datasets as shown in Table 2.2 were used.

The Forest/Non-Forest (FNF) [45] map, developed by the Japan Aerospace Exploration Agency (JAXA) provides a global representation of forest extent at a 25-meter resolution. The FNF map is derived from the Phased Array type L-band Synthetic Aperture Radar (PALSAR) on the Advanced Land Observing Satellite-2 (ALOS-2). In the first release of FNF maps in 2010, a threshold-based method was used to determine forest/non-forest for each year. Average accuracies over the four years (2007–2010) of mapping were 91.25%, 84.86% and 94.81% against different reference data [45]. In addition, starting from FNF Ver.2.0.0, a shift from conventional threshold-based method to Random Forest classifier adopted on PALSAR-2 data was done to improve the accuracy. Accuracy thus improved by 5-10 % or more compared to the previous version [29].

GHS-BUILT-S-R2023A [40] provides a global 10-meter resolution representation of total built-up surface for 2018. The Symbolic Machine Learning (SML) [39] method was used. A validation dataset by human visual interpretation was created to validate the product.

Dataset Name	Focus	Temporal Resolution	Spatial Res.
Forest/Non-Forest (FNF) V2	Forest	Annual (2007–2010, 2015–2024)	25 m
GHS-BUILT-S-R2023A	Built-up	2018	10 m
GHS-BUILT-S1	Built-up	2016	20 m
GHS-BUILT-LDSMT	Built-up	1990, 2000, 2014	38 m
Global Forest Cover (GFC)	Forest	2020	10 m
GISD30	Built-up	1985–2020 (5-year intervals)	30 m
Global Surface Water (GSW)	Water	Annual (1984–2021)	30 m
Global Urban Footprint (GUF)	Built-up	2011	12 m
GWL_FCS30D	Wetland	Annual (2000–2022)	30 m
NESEA-Rice10	Cropland	Annual (2017–2019)	10 m
Tree canopy cover	Forest	2000	30 m
World Settlement Footprint (WSF)	Built-up	2015, 2019	10 m

Table 2.2: Binary HRLC Datasets

The results indicate an Intersection over Union (IoU) of 0.92 for distinguishing between built-up and non-built-up [40].

GHS-BUILT-S1 [13] is a global, binary built-up layer developed by the Joint Research Centre (JRC) of the European Commission for 2016 derived from 5026 scenes of a large Sentinel-1A GRD collection in 2016. The generation of the product was based on the SML method. The learning data were extracted from FROM-GLC30 and GHS_BUILT_LDSMT v.2015. The output of SML provides as a confidence measure the evidence-based normalized differential index (ENDI). ENDI was then converted into a built-up/non-built-up mask via a threshold-based method. The Global Urban Footprint (GUF) was selected as the reference dataset for validation. Detailed results are not provided. The median Kappa varies in the range 0.25-0.5 [13].

GHS_BUILT_LDSMT v.2017 [13] is a global multi-temporal Landsat-derived built-up product, a sister product of GHS BU S1, which provides products in 1990, 2000 and 2014. The learning data were extracted from GlobeLand30 and GHS_BU_S1. The generation

of the product was also based on the SML method. The same validation strategy as the GHS_BU_S1's was applied in this case. Detailed results are not provided. The median Kappa varies in the range 0.25-0.5 which is comparable with GHS BU S1 [13].

The Global Forest Cover (GFC) [17] map developed by the JRC of the European Commission provides a global representation of forest presence for the year of 2020 at a 10-meter spatial resolution. GFC 2020 is generated through a synthesis approach that integrates multiple recent global datasets. The production workflow involves three main steps implemented in Google Earth Engine: first, determining the global maximum extent of tree cover; second, applying Boolean exclusion rules to mask out areas outside the forest definition; and third, post-processing. In terms of accuracy, a preliminary assessment using the IIASA global forest management reference dataset estimated an overall accuracy of 76.6%, with a user accuracy for the forest class of 95.2% [17].

GISD30 [51] developed by CAS is a global 30-m impervious surface dynamic dataset for 1985–2020 derived from time-series Landsat imagery on Google Earth Engine. It used a novel automated framework that combined automatic sample extraction with temporal spectral generalization to extract high-confidence samples and constrains mapping with a 2020 maximum impervious extents derived from multiple products. Then, random forest classifier was trained and spatiotemporal consistency correction was applied to improve temporal stability. In terms of accuracy assessment, validation with 23,322 time-series samples reports an overall accuracy of 90.1%, and the product encodes expansion timing in one band (pervious, pre-1985 impervious, and seven 5-year expansion intervals) [51].

Global Surface Water (GSW) [38] developed by JRC provides a globally consistent 30-m characterization of surface water distribution for 1984–2015. GSW was derived from all the archive of Landsat 5 TM, the Landsat 7 ETM+ and the Landsat 8 OLI images. Expert systems, visual analytics and evidential reasoning were exploited to do the mapping. In terms of accuracy assessment, validation with 40,124 globally distributed control points across sensors and years indicates <1% false water detections and <5% missed water overall [38].

The Global Urban Footprint (GUF) [16] developed by the German Aerospace Center (DLR) provides a global representation of built-up/non built-up based on the radar satellite imagery of the German satellites TerraSAR-X and TanDEM-X. A novel unsupervised and fully automated system based on advanced Support Vector Data Description (SVDD), Urban Footprint Processor (UFP), was proposed to handle more than 180,000 SAR images acquired between 2010 and 2013. In terms of accuracy assessment, the result indicates an overall accuracy of 86.6% [16].

GWL_FCS30D [52] developed by CAS provides the first global 30-m annual wetland maps with a fine classification system for 2000–2022, generated from time-series Landsat imagery on the Google Earth Engine platform. A random forest classifier was trained on a globally distributed training sample pool generated from more than 10 wetland-related products. In terms of accuracy assessment, a total of 22719 globally distributed validation points were collected for 2020, including 10952 non-wetland points and 11767 wetland points. The result indicates an overall accuracy of 86.95% [52]. Two third-party time-series validation datasets, Land Cover Monitoring Assessment and Projection (LCMAP) from the United States and Land Use/Cover Area frame Survey (LUCAS) from the European Union were used to analyze the accuracy variation during 2000–2018. The mean Producer Accuracy (PA) and User Accuracy (UA) for LCMAP and LUCAS are 78.32% and 93.52%, and 89.95% and 86.32% respectively [52].

NESEA-Rice10 [27] developed by Beijing Normal University (BNU) provides 10-meter annual paddy rice maps for Northeast and Southeast Asia from 2017 to 2019 based on time-series Sentinel-1 SAR and MODIS observations. Paddy rice is identified using a rule-based phenology approach centered on the transplanting-stage flooding signal. Validation relies on consistency checks against subnational agricultural statistics and existing rice products, reporting $R^2 = 0.80\text{--}0.97$ versus official statistics [27].

Tree canopy cover [28] developed by Hansen/UMD/ Google/USGS/NASA provides 30-meter resolution global maps of tree cover extent for 2000 using Landsat observations. Detailed methodology and accuracy assessment are not provided.

The World Settlement Footprint (WSF) [34, 35] developed by the DLR is a 10-meter resolution binary mask outlining the extent of human settlements globally. For the generation of WSF2015 [34], 2014–2015 Sentinel-1 SAR and Landsat-8 observations were utilized to train a RBF-SVM classifier. In terms of accuracy assessment, A large-scale validation based on about 900,000 crowdsourced photo-interpretation labels from Google Earth images shows that WSF2015 achieved an overall accuracy of 86% [34]. For the generation of WSF2019, 6 features of Sentinel-1 images and 25 features of Sentinel-2 images were utilized to train a random forest classifier. To assess the accuracy and reliability of the WSF2019 [35], two independent crowd-sourcing-based validation exercises have been carried out. Nevertheless, the details of accuracy of WSF2019 are not provided now.

2.3. Google Satellite Embedding

Google Satellite Embedding is a global collection of learned geospatial embedding which was produced by AlphaEarth Foundations (AEF) [7].

AEF is a geospatial embedding model trained on a variety of Earth observation (EO) datasets (e.g. optical satellite images, radar, 3D laser mapping and climate simulations). Instead of focusing on a particular task, the AEF model seeks to learn transferable, general-purpose representations of the Earth’s surface. The AEF introduces a foundation geospatial embedding field model designed to address the critical challenge of sparse high-quality labels [7]. Evaluation in realistic data-scarce scenarios was conducted by comparing the performance with other baseline models using pruned existing datasets. It was reported that AEF consistently outperforms designed and learned featurization approaches in relatively sparse data regimes [7].

Global annual embedding-field layers (Google Satellite Embedding) which have 64 dimensions at a 10-meter resolution for 2017–2024 are released on Google Earth Engine (GEE) now. Additionally, the embedding product is ready-to-use, which means no preprocessing of this product will be needed. Nevertheless, it is impossible to explain the meaning of each dimension because the 64 dimensions are learned by a deep learning model.

2.4. External Validation Datasets

In this study, an external validation dataset was used to improve the reliability of the validation results since self-established validation set can be biased.

SRS_val SRS_val is a novel stratified random sampling GLC validation dataset created by T.Zhao et al. (2023) [54] to assess the accuracies of 6 existing HRLCs. SRS_val contains 79,112 globally distributed samples for the year 2020. SRS_val is designed using a stratified equal-area random sampling scheme that explicitly increases sampling density in heterogeneous landscapes and rare land cover classes. In addition, the samples were labeled by manual interpretation with quality control.

2.5. Data Acquisition

The data providers and corresponding access portals for all evaluated land cover products are detailed in Table 2.3. With the exception of the Global Urban Footprint (GUF), all datasets were acquired through publicly accessible open-source repositories. The GUF dataset was specifically obtained via a private data-sharing link provided by the German Aerospace Center (DLR) team.

The following are the preprocessing procedures for all the implementations in this study. All the multi-class HRLC products were downloaded to cover the whole study area.

For the Dynamic World dataset, annual composites were generated by calculating the mode value from all available maps within a given year. The Dynamic World 2015 composite was derived using data from June 27, 2015, to June 30, 2016. This extended timeframe was necessary to compensate for the poor coverage in 2015, as the Dynamic World dataset only became available on June 27, 2015.

GLC_FCS30D was previously extracted from the provided multi-band product in which every band represents a land cover map for a single year.

Esri LULC was provided in the Universal Transverse Mercator (UTM) projection system and the fragments belong to different UTM zones. The fragments from Esri LULC were reprojected to World Geodetic System 84 (EPSG:4326) maintaining the equivalent spatial resolution of the originals to facilitate analysis and accelerate the generation of products.

GHS-BUILT-S1 2016 and GHS-BUILT-LDSMT 2014 were provided on a global scale. Therefore, they were clipped to the extent of the study area.

The metadata of "NoData" was updated with the correct value since some producers did not fill the "NoData" metadata (e.g., GlobeLand30).

Dataset Name	Provider	Access Point
<i>Multi-class HRLC Datasets</i>		
Dynamic World	Google / World Resources Institute	Link
Esri LULC	Esri / Microsoft / Impact Observatory	Link
FROM-GLC10	Tsinghua University	Link
FROM-GLC30	Tsinghua University	Link
GlobeLand30	NGCC	Link
GLC_FCS30D	CAS	Link
WorldCover	ESA	Link
<i>Binary HRLC Datasets</i>		
Forest/Non-Forest	JAXA	Link
GHS-BUILT-S-R2023A	JRC	Link
GHS-BUILT-S1	JRC	Link
GHS-BUILT-LDSMT	JRC	Link
Global Forest Cover	JRC	Link
GISD30	CAS	Link
Global Surface Water	JRC	Link
Global Urban Footprint	DLR	Link
GWL_FCS30D	CAS	Link
NESEA-Rice10	BNU	Link
Tree canopy cover	Hansen/UMD/ Google/USGS/NASA	Link
WSF	DLR	Link
Google Satellite Embedding	Google	Link
<i>External Validation Datasets</i>		
SRS_val	CAS	Link

Note: All websites provided in the Access Point column were last accessed on January 29, 2026.

Table 2.3: Accessibility of Datasets

2.6. Legend Harmonization

Legend harmonization means matching those legends that have almost identical meaning. To overcome the limitations of compatibility and comparability among all the products provided by different institutes, legend harmonization for the classification systems is a necessity. The legend harmonization in this study was done with reference to the CCI HRLC legend correspondence document (Q. Xu, Personal Communication, January 29, 2026). The legend matching was performed based on the definitions of each class of all the HRLC datasets.

First, the class names and descriptions of all the used HRLC datasets were collected. Subsequently, a preliminary matching between the target classes and source classes was manually done based on the descriptions of each class; the classes that cannot be mapped to the target classes were discarded and marked as "not used." Then, domain experts reviewed the preliminary correspondence, identifying potential mistakes. Discrepancies identified during the review were resolved to finalize the harmonization rules.

The correspondence between the legend of the land cover classification system used in this study, namely MOLCA legend [3] and FAO LCCS [19] is shown in Table 2.4. FAO LCCS defines land cover classes with a string of classifiers (Presence/Absence of Vegetation, Edaphic Condition, and Artificiality of Cover).

Table 2.5 shows the detailed rule of the legend harmonization in this study. Figure 2.2 illustrates the mapping relationship between different source datasets and the MOLCA legend system. This matrix visually demonstrates into which MOLCA classes each HRLC product and validation dataset are reclassified.

MOLCA Code	MOLCA Class	LCCS Code	Presence of Vegetation	Edaphic Condition	Artificiality of Cover
8	Cropland	A11	Primarily vegetated	Terrestrial	Artificial/managed
20	Forest	A11, A12	Primarily vegetated	Terrestrial	(Semi-)Natural, and Artifi- cial/Managed
7	Grassland	A11, A12	Primarily vegetated	Terrestrial	(Semi-)Natural, and Artifi- cial/Managed
5	Shrubland	A11, A12	Primarily vegetated	Terrestrial	(Semi-)Natural, and Artifi- cial/Managed
9	Wetland	A23, A24	Primarily vegetated	Aquatic or Regularly Flooded	(Semi-)Natural, and Artifi- cial/Managed
13	Built-up	B15	Primarily non-vegetated	Terrestrial	Artificial/Managed
12	Bareland	B16	Primarily non-vegetated	Terrestrial	(Semi-)Natural
16	Permanent ice and snow	B27, B28	Primarily non-vegetated	Aquatic or Regularly Flooded	(Semi-)Natural, and Artifi- cial/Managed
15	Water	B27, B28	Primarily non-vegetated	Aquatic or Regularly Flooded	(Semi-)Natural, and Artifi- cial/Managed

Table 2.4: Mapping of MOLCA Classes to FAO LCCS

Table 2.5: Legend Harmonization between MOLCA and Existing Datasets

Dataset	Forest	Shrubland	Grassland	Cropland	Wetland	Lichens/Moss	Bareland	Built-up	Water	Ice/Snow	Not Used
<i>Multi-class Datasets</i>											
GLC_FCS30D	52, 62, 72, 82	120–122	130	10, 11, 20	181–187	140	200–202, 150, 152, 153	190	210	220	12, 51, 61, 71, 81, 91, 92
Dynamic World	1	-	2	4	3	-	7	-	0	8	5, 6
LULC Annual v2	2	-	-	5	4	-	8	7	1	9	10, 11
WorldCover	10	-	30	40	90, 95	100	60	50	80	70	20
GlobeLand30	20	-	30	-	50	70	90	80	60	100	10, 40
FROM-GLC10 2017	20	40	30	10	50	-	90	80	60	100	70
FROM-GLC30 15/17	2	4	3	1	5	-	9	8	6	10	7
FROM-GLC30 2010	21, 22	40	30–32	10, 11, 13, 94	50, 51	-	52, 90–93, 96	80–82, 12	60–64	100–102	20, 23, 24, 70–72, 95
<i>Binary Datasets</i>											
GISD30	-	-	-	-	-	-	-	1 (<1985); 1,2 ('85-'90); 1...3 ('90-'95); 1...4 ('95-'00); 1...5 ('00-'05); 1...6 ('05-'10); 1...7 ('10-'15); 1...8 ('15-'20)	-	-	0
WSF	-	-	-	-	-	-	-	255	-	-	0
GUF	-	-	-	-	-	-	-	255	-	-	0
GHS-BUILT-S	-	-	-	-	-	-	-	[50, 100]	-	-	[0, 50]
GHS-BU-S1	-	-	-	-	-	-	-	1	-	-	-
GHS-BU-LDSMT	-	-	-	-	-	-	-	101	-	-	1
GSW	-	-	-	-	-	-	-	-	2, 3	-	1
GWL_FCS30D	-	-	-	-	181–187	-	-	-	180	-	0
GFC	1	-	-	-	-	-	-	-	-	-	0
Tree Canopy	[50, 100]	-	-	-	-	-	-	-	-	-	[0, 50]
FNF3	1	-	-	-	-	-	-	-	3	-	2
FNF4	1, 2	-	-	-	-	-	-	-	4	-	3
NESEA-Rice10	-	-	-	1	-	-	-	-	-	-	-
<i>Validation Datasets</i>											
SRS_val	50, 60, 70, 80	120	130	10, 20	180	140	200	190	210	220	90

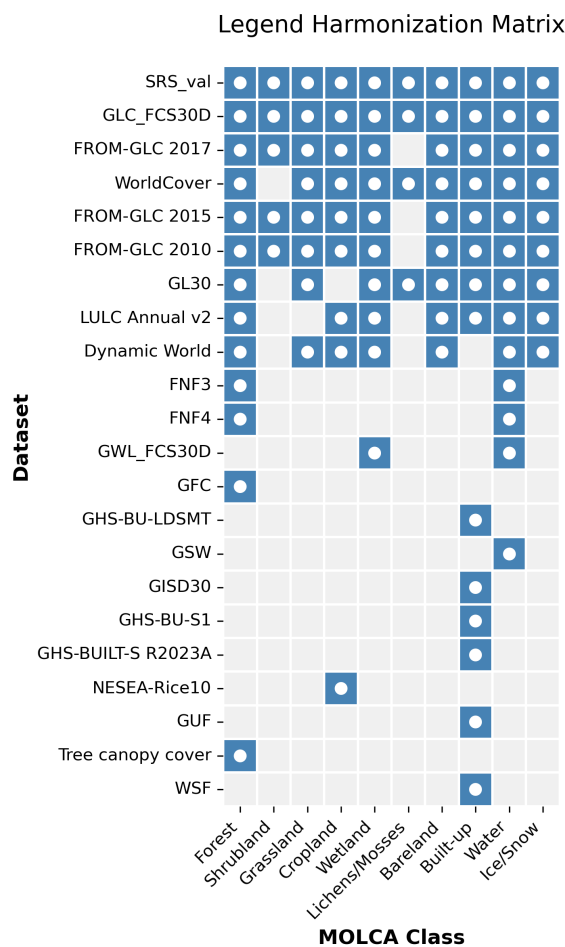


Figure 2.2: Mapping Matrix of Legend Harmonization

2.7. Critical Analysis

Although the aforementioned global land cover products and validation datasets provide global coverage, their consistency and mapping accuracy vary significantly across different parts of the Earth. Considering the products generated in this study highly rely on the quality of the existing HRLC products, some critical analyses were conducted to understand the global and local consistency of HRLC products and validation datasets. The following section presents a localized data analysis, examining the spatial distribution of the external validation dataset and performing a comparative accuracy assessment of the existing LC maps.

2.7.1. Distribution of External Validation Datasets

Given that the external validation dataset SRS_val is provided on a global scale, it was necessary to assess the suitability for the study area. The spatial and class count distributions of it were analyzed to ensure the dataset could provide effective validation.

SRS_val SRS_val was provided in .csv format with the information of latitude, longitude, and class label. The latitude and longitude were treated as referenced to WGS 84 datum in this study. The sample points were previously clipped to the boundary of the study area. Subsequently, the sample points were relabeled according to the legend harmonization rules shown in Table 2.5 to ensure consistency.

Figure 2.3 and Figure 2.4 show the class count distribution the spatial distribution of SRS_val.

A total of 372 validation sample points are widely distributed across the study area. Nevertheless, the dataset shows an imbalance of classes. Classes "Permanent ice and snow" and "Lichen and Mosses" do not appear in the study area due to actual situation, thus 8 classes exist in the study area. The "Forest" class is the most prevalent, containing 213 samples, which is significantly higher than any other class. The classes with the lowest representation are "Built-up" (7 samples), "Wetland" (7 samples), and "Bareland" (2 samples).

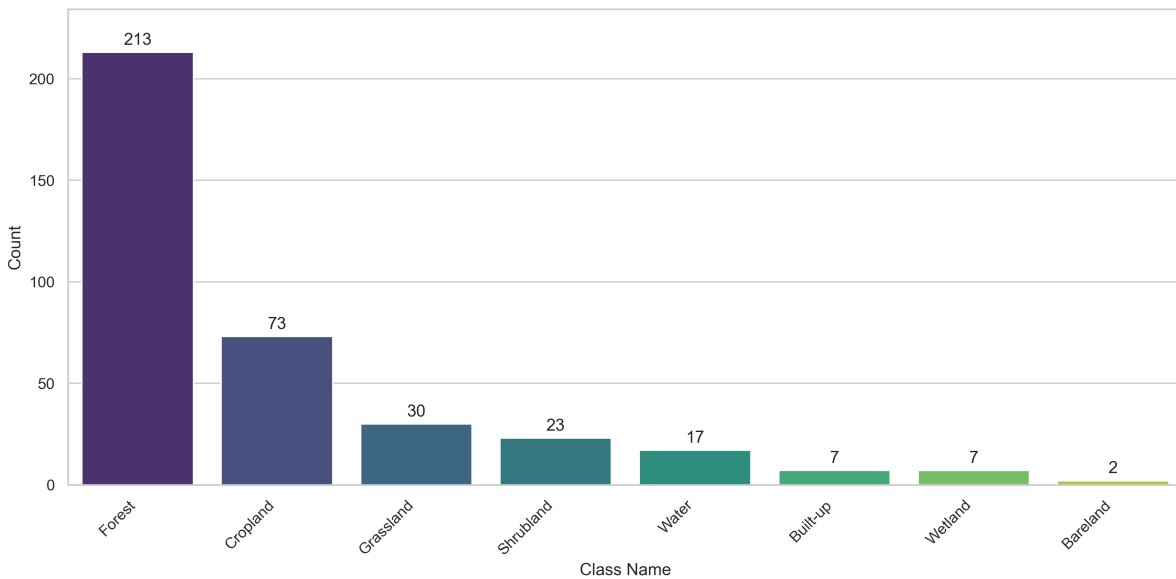


Figure 2.3: Class Count Distribution of SRS_val

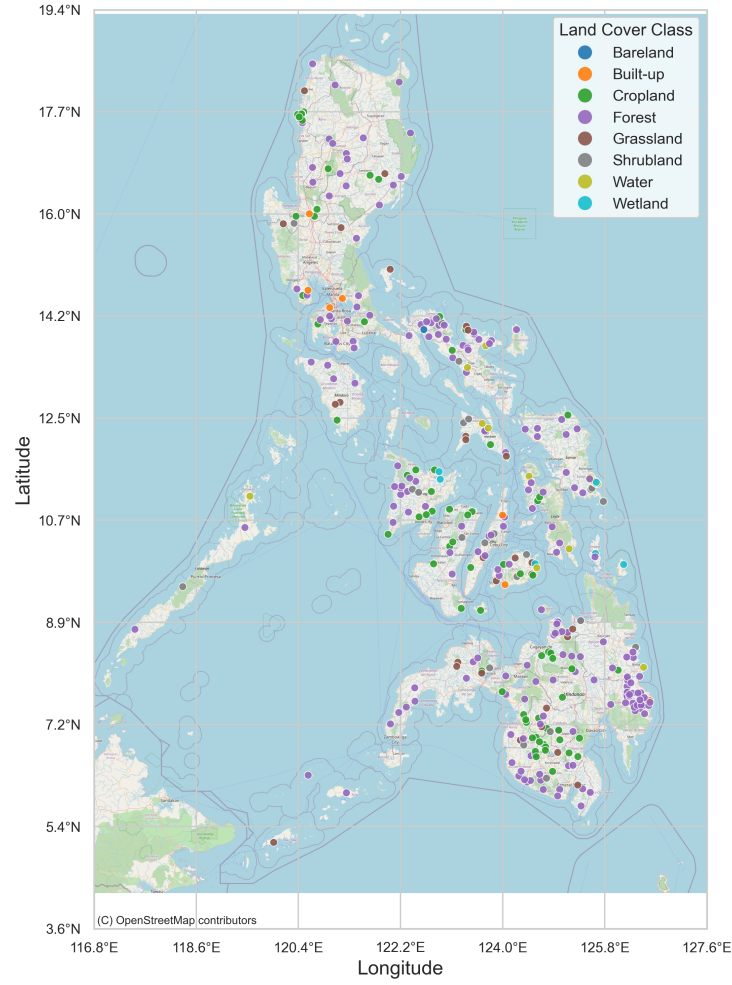


Figure 2.4: Spatial Distribution of SRS_val

2.7.2. Analysis on Multi-class HRLCs

This section includes the statistical analysis of the multi-class HRLC dataset and the accuracy assessment with SRS_val. The year 2019 was selected as the baseline year because it is one of the target years in this study and all multi-class HRLC datasets are available for the year. Since some products are not available for the year 2019, the products with the closest baseline year were used as alternatives.

Datasets to be analyzed: Dynamic World 2019, Esri LULC 2019, FROM-GLC10 2017, FROM-GLC30 2017, GLC_FCS30D 2019, GlobeLand30 2019, WorldCover 2020.

The HRLC datasets were obtained and preprocessed as described in Section 2.5. Since the HRLC datasets were typically provided in fragments, fragments belonging to the same datasets were merged. To make the pixel counts comparison more intuitive, HRLCs at

10-meter and 30-meter resolution were further resampled with reference to FROM-GLC10 and GLC_FCS30D respectively.

2.7.2.1. Statistical Analysis

Conducting a statistical analysis of HRLC class distribution is essential for several reasons: it provides a comprehensive overview of land cover composition to inform experimental design, enables cross-dataset comparisons, and significantly enhances the interpretability of the results of experiments in this study.

Class distribution of each HRLC map was calculated individually, then they are mapped according to the abovementioned legend harmonization rules. The "Not Used" (pixels that can not be mapped to any MOLCA code) and "No Data" pixels were also included in the statistics to make the percentage more intuitive, ensuring that the derived percentages accurately reflect the entire geographical extent of the datasets rather than just the mapped areas.

Figure 2.5 shows an intuitive comparison of land cover distribution of different multi-class HRLC products. And Tables 2.6 and 2.7 provide the pixel counts and percentage distributions for each class and dataset.

It can be easily observed that "Forest" is the predominant class followed by "Cropland," "Grassland" and "Water," which can be defined as "major classes." Other classes can be defined as "minor classes." The class "Lichens and Mosses" does not appear on any of the HRLC maps. Similarly, the class "Permanent ice and snow" only appears on two HRLC maps with a very small proportion. Therefore, these two classes can be considered practically non-existent.

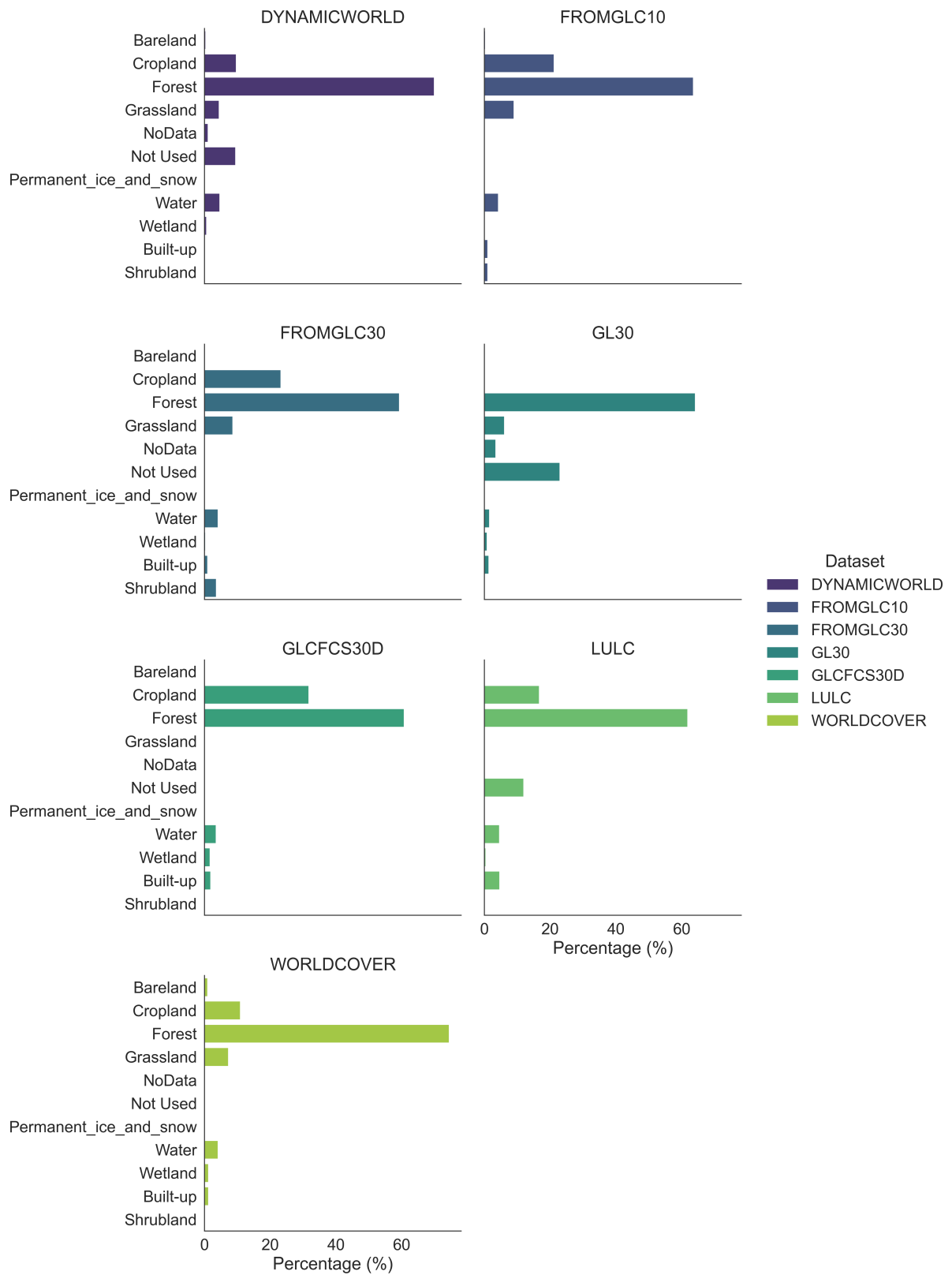


Figure 2.5: Class Distribution for each Dataset

Class	Dynamic World	FROM-GLC10	LULC	WorldCover	FROM-GLC30	GL30	GLC_FCS30D
Bareland	9,583,203	6,885,364	5,189,031	27,959,500	303,346	269,047	736,569
Built-up	0	28,111,789	136,862,317	34,793,241	3,168,306	4,252,219	6,205,759
Cropland	288,454,619	632,059,433	498,164,808	325,334,403	77,124,656	0	105,400,216
Forest	2,091,707,981	1,900,037,807	1,849,496,465	2,226,879,480	196,999,288	213,210,249	201,864,071
Grassland	131,295,516	268,566,983	0	217,810,713	28,524,732	19,984,776	1,138
Permanent ice/snow	873,407	0	138	0	0	0	0
Shrubland	0	29,395,852	0	0	11,906,126	0	708,368
Water	138,497,962	125,777,845	135,880,931	122,708,223	13,611,027	4,945,959	11,593,897
Wetland	17,219,893	1,946,693	11,082,175	35,725,153	840,915	2,478,906	5,597,033
NoData	32,614,528	0	76	0	53,093	11,262,882	6,118
Not Used	282,534,657	0	356,105,716	1,570,944	0	76,127,451	418,320
Total Pixels	2,992,781,766	2,992,781,766	2,992,781,657	2,992,781,657	332,531,489	332,531,489	332,531,489

Table 2.6: Pixel Count Distribution by Class

Class	Dynamic World	FROM-GLC10	LULC	WorldCover	FROM-GLC30	GL30	GLC_FCS30D
Bareland	0.3%	0.2%	0.2%	0.9%	0.1%	0.1%	0.2%
Built-up	0.0%	0.9%	4.6%	1.2%	1.0%	1.3%	1.9%
Cropland	9.6%	21.1%	16.6%	10.9%	23.2%	0.0%	31.7%
Forest	69.9%	63.5%	61.8%	74.4%	59.2%	64.1%	60.7%
Grassland	4.4%	9.0%	0.0%	7.3%	8.6%	6.0%	<0.1%
Permanent ice/snow	<0.1%	0.0%	<0.1%	0.0%	0.0%	0.0%	0.0%
Shrubland	0.0%	1.0%	0.0%	0.0%	3.6%	0.0%	0.2%
Water	4.6%	4.2%	4.5%	4.1%	4.1%	1.5%	3.5%
Wetland	0.6%	0.1%	0.4%	1.2%	0.3%	0.7%	1.7%
NoData	1.1%	0.0%	<0.1%	0.0%	<0.1%	3.4%	<0.1%
Not Used	9.4%	0.0%	11.9%	0.1%	0.0%	22.9%	0.1%

Table 2.7: Percentage Distribution by Class

2.7.2.2. Accuracy Assessment

In order to have an overview of all the multi-class HRLC datasets and to understand whether they are suitable enough to be used in the study area, the accuracy assessment on seven HRLC datasets were conducted before entering the generation phase. The assessment also enables inter-comparison between the existing land cover maps and the land cover maps generated in this study.

Previous study by T.Zhao et al. (2023) [54] used a created validation dataset (SRS_val) to assess the global and regional accuracy and consistency of six multi-class HRLC products which include 5 out of 7 multi-class HRLC products used in this study except for Dynamic World and GLC_FCS30D. The samples distributed in the study area were selected to assess the accuracy for the year 2019 which is one of the target years in this study and

close to the baseline year of SRS_val (2020). All seven abovementioned HRLC products for 2019 were assessed with SRS_val to have a comprehensive understanding.

Figure 2.6 shows the flow of the accuracy assessment.

The validation samples were clipped to the boundary of the study area and reclassified according to the legend harmonization rules. The coordinate transformation for the sample points would be performed if the CRS of the assessed HRLC map was detected different from WGS84 (EPSG:4326) since the provided coordinates of the sample points are treated as referenced to WGS84 datum. Points from the HRLC maps were sampled according to the validation points' coordinates and the corresponding class codes of the points were mapped according to the legend harmonization rule. If the sample points from HRLC represents NoData or the class code could not be mapped, the points would be considered invalid. Then invalid sample points were removed. In the end, the confusion matrices were built and the metrics were computed.

As shown in Table 2.8, the Overall Accuracies (O.A.) of six HRLC datasets were evaluated against the SRS_val within the study area and compared with their reported global and Asian regional accuracies. Esri LULC achieved the highest accuracy in the study area at 83.39%, significantly outperforming its reported global (58.9%) and Asian (59.61%) validation scores.

Nevertheless, the recall and precision for each class as shown in Table 2.9 reveal the O.A. should be interpreted with caution. The O.A. is primarily driven by the product's robust performance in "Forest" and "Water," which constitute a significant portion of the study area. Conversely, the poor performance can be seen in "Shrubland" and "Bareland." Across all 7 HRLC products, the precision and recall for both classes are zero except for "Shrubland" of From-GLC30 2017, which probably results from the lack of samples in the study area or the quality issue of samples for specific classes.

In conclusion, the multi-class HRLC datasets shows high consistency of overall accuracy on different scales. The overall accuracy in the study area is always close to the claimed one. The result also reflects the excellent validity of SRS_val in the study area. Therefore, all HRLC datasets are considered valid in this study. But obviously, all of them are not competent enough to be a benchmark land cover map.

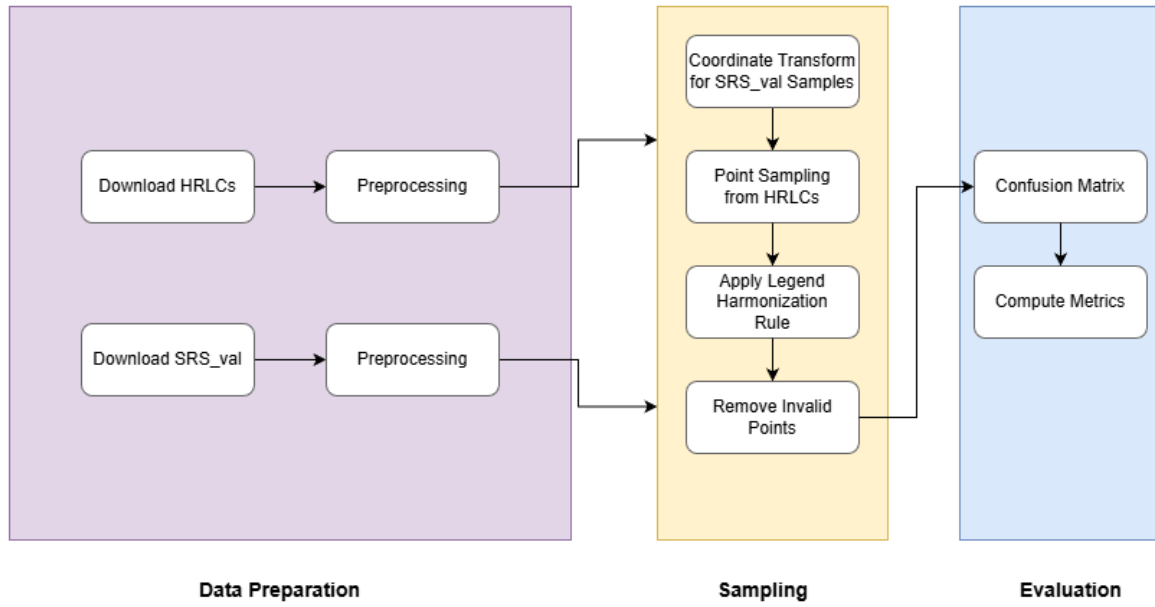


Figure 2.6: Flow of Accuracy Assessment

HRLC Dataset	Claimed O.A.	SRS_val O.A. (Global) [54]	SRS_val O.A. (Asia) [54]	SRS_val O.A. (Study Area)
Dynamic World	73.80%	—	—	76.95%
Esri LULC	85.00%	58.90%	59.61%	83.39%
FROM-GLC10	72.76%	68.98%	73.59%	69.62%
FROM-GLC30	64.90%	66.30%	69.24%	72.85%
GLC_FCS30D	80.88%	—	—	75.54%
GlobeLand30	80.33%	69.96%	75.10%	76.26%
WorldCover	74.40%	70.54%	66.80%	73.39%

Table 2.8: Comparison of Overall Accuracy (O.A.) for HRLC Datasets

Class	FROM-GLC10		FROM-GLC30		Dynamic World		GL30		GLC_FCS30D		Esri LULC		WorldCover	
	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
Shrubland	0.00	0.00	0.27	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Grassland	0.32	0.40	0.30	0.30	0.39	0.41	0.40	0.42	0.00	0.00	0.00	0.00	0.39	0.40
Cropland	0.52	0.56	0.55	0.71	0.92	0.52	0.00	0.00	0.54	0.90	0.87	0.77	0.85	0.45
Wetland	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.33	0.60	0.86	0.00	0.00	0.25	0.14
Bareland	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Built-up	1.00	0.29	1.00	0.43	0.00	0.00	0.67	0.67	0.67	0.57	0.47	1.00	1.00	0.57
Water	0.77	1.00	0.71	1.00	0.74	1.00	0.33	1.00	0.89	0.94	0.74	1.00	0.77	1.00
Forest	0.82	0.88	0.91	0.87	0.79	0.97	0.81	0.94	0.88	0.89	0.86	0.97	0.76	0.97

Table 2.9: Comparison of Precision and Recall for Different Land Cover Products

3 | Methodology

3.1. Overview

Figure 3.1 illustrates the overall workflow of this study, which can be divided mainly into three phases.

The initial phase focused on downloading and harmonizing the input datasets. It involved the existing Multi-class HRLC Maps and Binary HRLC Maps. A critical step in this phase was legend harmonization as illustrated in Section 2.6, in which the various classification schemes of the input maps were harmonized into a common target legend to ensure interoperability.

The second phase was the generation of the Map of Land Cover Agreement (MOLCA) and Map of Land Cover Partial Agreement (MOLCOPA) with input datasets. MOLCA was generated on the basis of a unanimous consensus approach, which led to some limitations. Thus, a probabilistic consensus approach was proposed, which produced MOLCOPA, a count map and various land cover maps. Additionally, MOLCA can also be converted from MOLCOPA and the count map.

Land cover maps with constrained coverage were produced with MOLCA and MOLCOPA. MOLCA can be regarded as a benchmark land cover map and MOLCOPA can provide extra information. The release of Google Satellite Embedding also enables the GeoAI-based land cover mapping.

In the third phase, land cover mapping was realized by combining GeoAI-based methods. The GeoAI-based methods fuse Google Satellite Embeddings with the consensus information from MOLCA and MOLCOPA. Thus, a complete land cover map could be produced.

In addition, evaluations were performed on the generated products in this study.

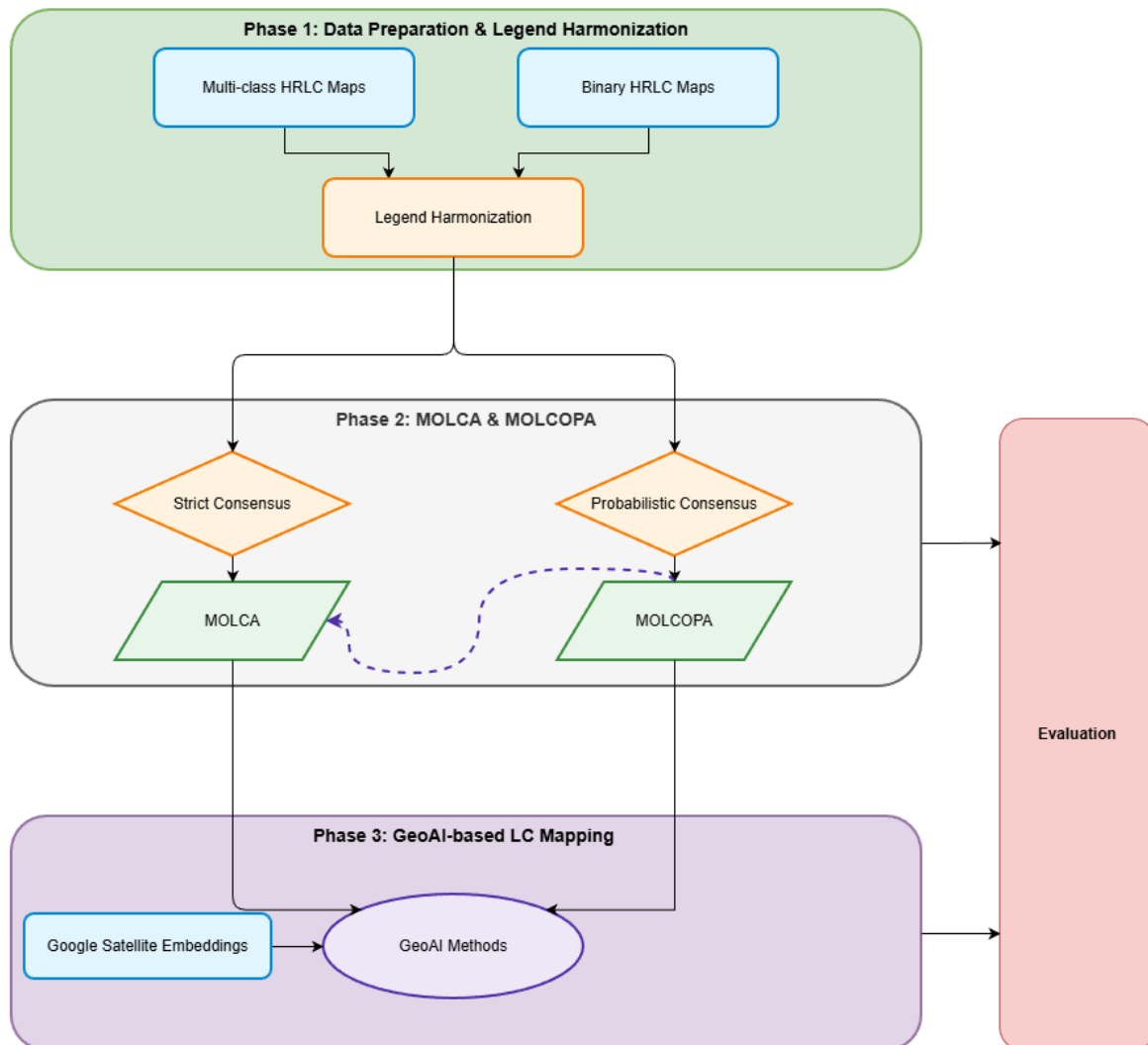


Figure 3.1: Overall Workflow

3.2. Introduction to Techniques

In this section, a brief introduction to the techniques used in this study is included.

3.2.1. Tiling System

The tiling system used in this study is identical to the tiling system of Sentinel-2 products, namely a slightly modified U.S. Military Grid Reference System (MGRS) [36].

The U.S. Military Grid Reference System [26] was designed for use with Universal Transverse Mercator (UTM) and Universal Polar Stereographic (UPS) grids. The world is generally divided into 6° by 8° geographic areas, each of which is given a unique identification, called the Grid Zone Designation (GZD). The GZD is subdivided into a collection of 100-km squares. Each square is identified by two letters, called the 100-km square identification, which is unique within the area covered by the GZD.

The tiling system used for Sentinel-2 products is a modified version of MGRS. The tiling grid consists of MGRS squares expanded from $100 \text{ km} \times 100 \text{ km}$ to $109.8 \text{ km} \times 109.8 \text{ km}$ [1].

The file of the tiling system was obtained from SentiWiki (Last Accessed 27 January 2026). Although the Sentinel-2 User Handbook [18] claims that the tile size is $100 \text{ km} \times 100 \text{ km}$, the file proves that the size is $109.8 \text{ km} \times 109.8 \text{ km}$ as reported by B. Bauer-Marschallinger et al. (2023) [1] and National Aeronautics and Space Administration (NASA) [36].

The provided file was in Keyhole Markup Language (KML) format and was converted to a GeoPackage (GPKG) file by extracting the embedded geometry information. In addition, the corresponding UTM zones of the tiles were also extracted. Subsequently, the tiles were also rasterized at 10-m and 30-m resolution with their corresponding UTM projection. Then they were reprojected to the target coordinate reference system World Geodetic System 1984 (EPSG: 4326), thus maintaining the equivalent spatial resolution for the convenience of generating products later.

Figure 3.2 shows the tiling grid used in this study. It was manually selected to cover the whole study area, which consists of 105 tiles in total.

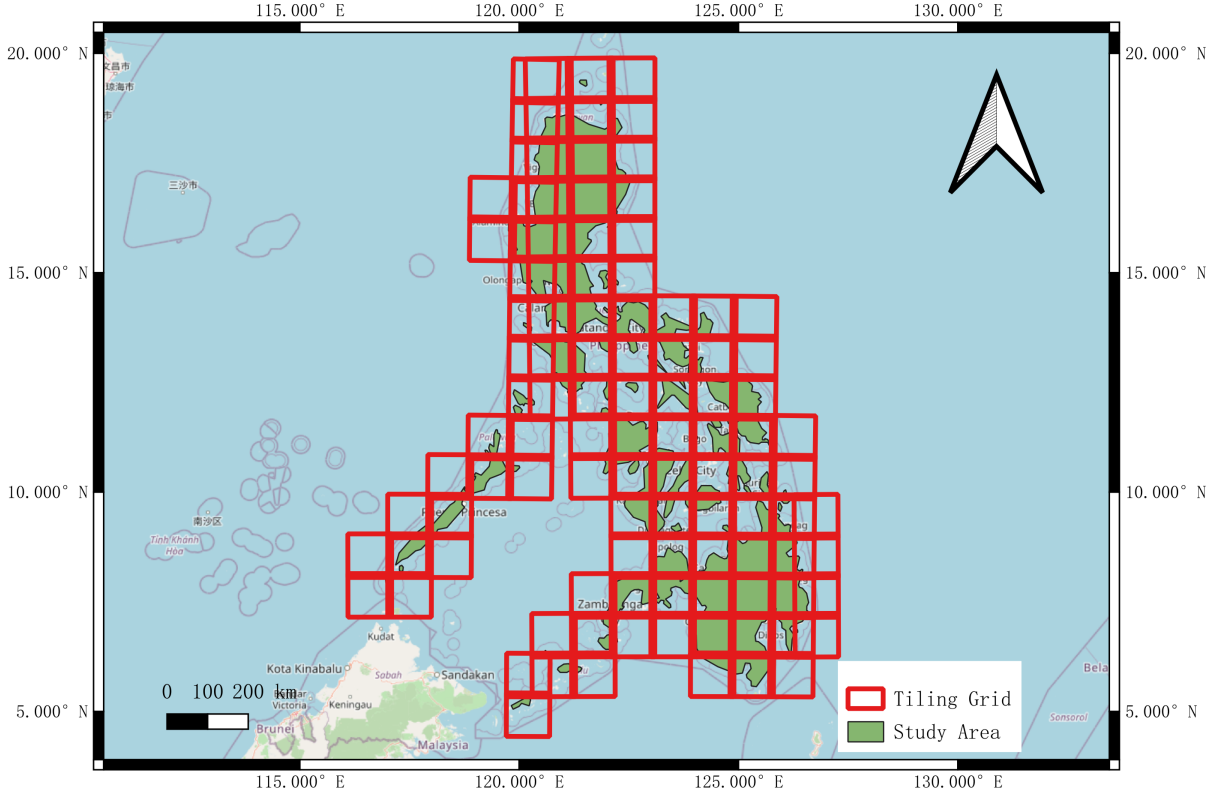


Figure 3.2: Tiling Grid

3.2.2. Machine Learning Algorithm

K-means K-means [33] is an unsupervised learning algorithm that partitions a dataset into K distinct clusters by minimizing the Within-Cluster Sum of Squares (WCSS).

Random Forest Random Forest (RF) [5] is an ensemble learning method that constructs a multitude of decision trees during training and outputs the class that is the mode of the classes of the individual trees.

Support Vector Machine Support Vector Machine (SVM) [14] is a supervised learning algorithm that finds the optimal decision boundary, known as the hyperplane, to maximize the margin between classes.

Multilayer Perceptron The Multi-Layer Perceptron [42] is a foundational feedforward artificial neural network (ANN) consisting of an input layer, hidden layers with non-linear

activation functions, and an output layer. MLP is able to approximate a wide range of nonlinear functions, a property guaranteed by the universal approximation theorem [48].

3.2.3. Other Techniques

SMOTE Synthetic Minority Over-sampling Technique (SMOTE) [9] is an oversampling method used to address class imbalance in datasets by generating synthetic minority class examples through interpolation between existing instances.

Cochran's Formula Cochran's formula [12] is a technique to determine the optimal sample size. The formula ensures that the sample is statistically representative of the population at a specified confidence level and margin of error. The equation is defined as Equation 3.1:

$$n = \frac{Z^2 p(1 - p)}{e^2} \quad (3.1)$$

Where:

- n represents the ideal sample size.
- Z is the Z-value. For a 95% confidence level, Z is typically set to 1.96.
- p is the estimated proportion of an attribute that is present in the population. To obtain the maximum sample size, p is set to 0.5.
- e is the margin of error.

3.3. Unanimous Consensus: MOLCA

MOLCA is a benchmark land cover map proposed to address the issue that obtaining sufficient training data for a global land cover map is time-consuming. MOLCA can easily provide numerous training data for a machine learning or deep learning task. The generation of MOLCA was based on the study by G. Bratic et al. (2023) [3] but slightly revised, which applied a unanimous consensus rule.

3.3.1. Overall Workflow

Figure 3.3 illustrates the overall workflow of the generation of MOLCA. The workflow is organized into four phases:

- (1) **Target Defining:** The process began with defining the targets, including the target years and spatial resolutions to ensure the consistency across datasets. In addition, the input datasets for the target years were determined.
- (2) **Data Preparation:** In this phase, all the data sets determined for the target years were prepared.
- (3) **Unanimous Consensus Rule:** The core mechanism was applied here. A strict consensus-based rule was applied, retaining only those pixels where all valid input maps were in total agreement, while marking discordant pixels as "NoData."
- (4) **Output & Evaluation:** The final output was the MOLCA, a conservative but high-confidence benchmark land cover map. Then MOLCA was evaluated to ensure reliability.

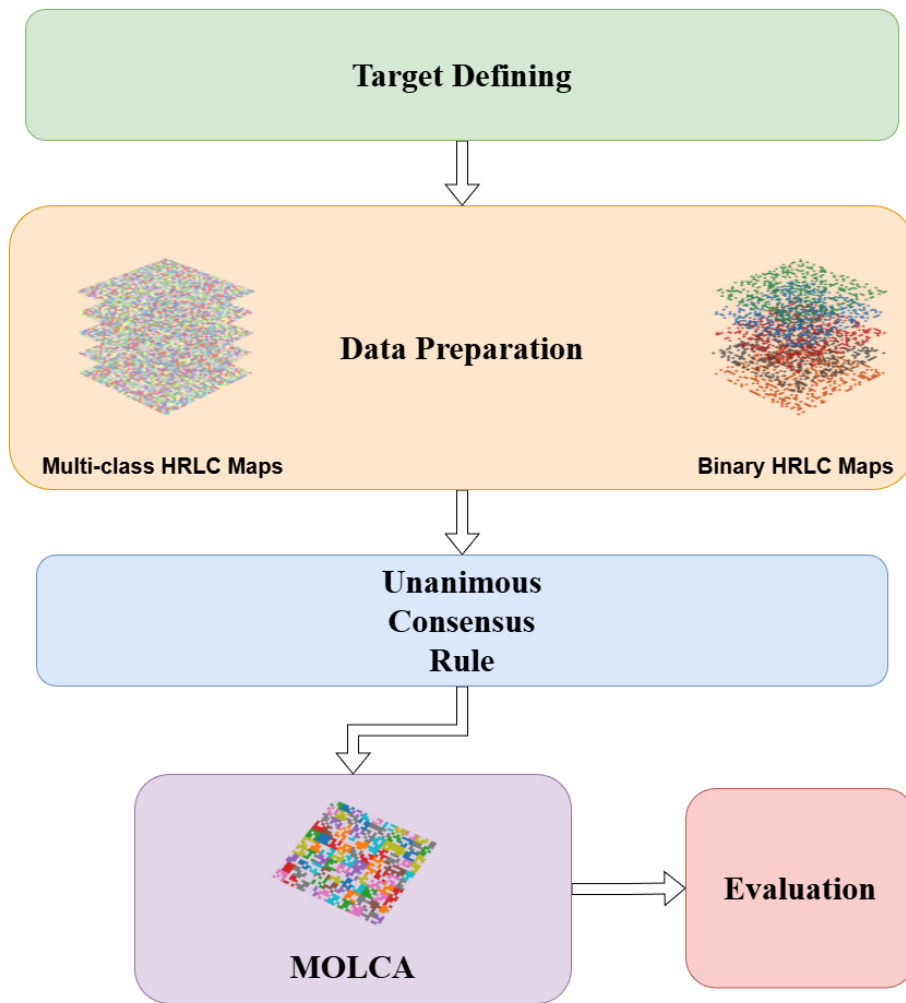


Figure 3.3: Overall Workflow of the MOLCA Generation

3.3.2. Target Defining and Data Preparation

Figure 3.4 shows the available epochs of the HRLC products mentioned in Chapter 2. A clear temporal trend is observable: the number of available HRLC products increases over the years.

To construct a long-term benchmark land cover dataset, initial temporal targets for MOLCA were established at five-year interval. The selection of these specific epochs was driven by the temporal availability of the input global HRLC products, as illustrated in Figure 3.4.

The year 2000 was identified as the earliest viable starting point. For the epochs before 2000, only few HRLC maps are available, which might result in a low-confidence MOLCA with a consensus-based method.

The year 2005 was excluded from the targets since there is only one available multi-class HRLC for 2005, which violated the rule defined in Section 3.3.3.

The year 2020 was replaced by the year 2019 based on the count of potential HRLC products for the epochs. Observing that some datasets are available for 2017 or 2018 but not for 2020, it was better to replace the target year 2020 with 2019.

The final selected target years were 2000, 2010, 2015 and 2019.

To maximize the reliability of the consensus-based product MOLCA, a diverse collection of HRLC products was aggregated. Table 3.1 details the specific datasets used for each target year (2000, 2010, 2015, and 2019).

Since HRLC products are not available for every target year, the closest available source year of the HRLC product was used as an alternative if it is not available for the target year. For example, FROM-GLC10 2017 was used to generate the MOLCA of 2019. In addition, FROM-GLC30 2017 was not used for the target year 2019 considering that FROM-GLC10 provides a better spatial resolution and has a matched performance compared to FROM-GLC30.

Furthermore, the spatial resolution of the final MOLCA product was determined by the dominant resolution of the available inputs. For 2000, 2010, 2015, the target spatial resolution is 30 meters. Considering that six multi-class HRLC products and eight binary HRLC products are available for 2019, including four 10-m multi-class and four 10-m binary HRLC products, a spatial resolution of 10 meters was adopted.

The datasets were obtained and preprocessed as described in Section 2.5.

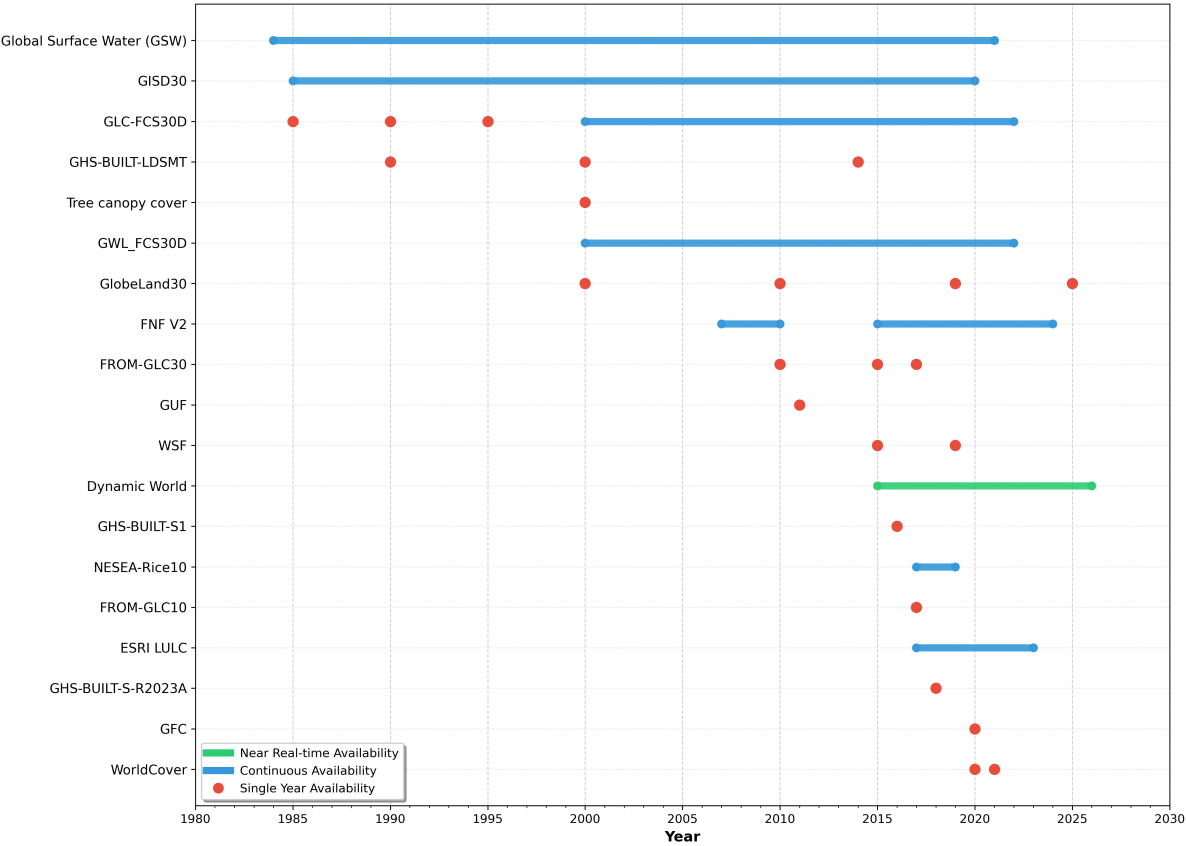


Figure 3.4: Temporal Availability of Datasets

All the used datasets were harmonized based on the harmonization rules as described in Section 2.6.

Dataset Name	Resolution	Source Year used for Target Year			
		2000	2010	2015	2019
<i>Multi-class HRLC Datasets</i>					
Dynamic World	10-m			2015	2019
Esri LULC	10-m				2019
FROM-GLC10	10-m				2017
FROM-GLC30	30-m		2010	2015	
GlobeLand30	30-m	2000	2010		2019
GLC_FCS30D	30-m	2000	2010	2015	2019
ESA WorldCover	10-m				2020
<i>Binary HRLC Datasets</i>					
FNF	25-m		2010	2015	2019
GHS-BUILT-S-R2023A	10-m				2018
GHS-BUILT-S1	20-m			2016	
GHS-BUILT-LDSMT	38-m	2000		2014	
GFC	10-m				2020
GISD30	30-m	2000	2010	2015	2019
GSW	30-m	2000	2010	2015	2019
GWL_FCS30D	30-m	2000	2010	2015	2019
GUF	12-m		2011		
NESEA-Rice10	10-m				2019
Tree Canopy Cover	30-m	2000			
WSF	10-m			2015	2019
Total Inputs	-	7	8	10	14
Output Resolution	-	30-m	30-m	30-m	10-m

Table 3.1: Summary of Target

3.3.3. Unanimous Consensus Rules

To generate the final consensus map, a rule defined by Equation (3.2) was applied. For a pixel p to be considered valid in MOLCA, it must satisfy two criteria simultaneously:

- (1) **Agreement among all Inputs:** Every input map that covers pixel p must report the exact same land cover class c . Any discrepancy among the inputs results in rejection ("NoData").
- (2) **Multi-class Support:** To ensure reliability, the pixel must be valid in at least two multi-class HRLC products.

$$\left\{ \begin{array}{l} MOLCA(p) = \begin{cases} c & \text{if } \exists c \in \text{Classes s.t.} \\ & \forall i \in \{1, \dots, K\}, M(i, p) = c \text{ (where valid)} \\ & \text{AND } N_{mc}(p) \geq 2 \\ \text{NoData} & \text{otherwise} \end{cases} \\ N_{mc}(p) = \sum_{i=1}^{K_{mc}} \mathbb{I}(M_{mc}(i, p) \neq \text{NoData}) \end{array} \right. \quad (3.2)$$

Where:

- $M(i, p)$ is the class value of the i -th input map at pixel p .
- $M_{mc}(i, p)$ is the class label of the i -th multi-class map at pixel p .
- $N_{mc}(p)$ is the count of valid (non-NoData) multi-class maps available at pixel p .
- K is the total number of input datasets.
- K_{mc} is the total number of multi-class input datasets.
- $\mathbb{I}(\cdot)$ is the indicator function.

3.3.4. Implementation

The proposed MOLCA generation workflow was implemented using GRASS GIS driven by Python scripting. The generation of MOLCA follows Algorithm 3.1. The processing was carried out tile-by-tile with the tiling system mentioned above.

First, data harmonization was performed for each tile. The tile was set as the region in GRASS GIS using `g.region`. Input datasets were imported using `r.import`. Raster fragments belonging to the same dataset were identified and mosaicked (using `r.patch`) to ensure spatial continuity. These layers were automatically aligned to the tile through reprojection and resampling with reference to the tile, followed by a reclassification step (using `r.reclass`) based on the legend harmonization rules to unify the legend.

Harmonized raster layers were stacked to compute the cross-product using `r.cross` of GRASS GIS, generating a new synthetic raster where each pixel value corresponds to a unique combination of class labels from the input maps (e.g., *Forest; Cropland; Forest*).

The attribute table of this cross-product map was then extracted using `r.category` and analyzed. For each pixel, the number of unique class values N_{unique} was calculated. The

consensus was considered reached only when $N_{unique} = 1$ for a pixel; thus, the agreed class C could be assigned to the pixel. Otherwise, the pixel would be marked as "NoData."

Finally, to ensure reliability, a Multi-class Constraint was applied. A validity count map (N_{mc}) was generated by stacking all valid multi-class layers (using `r.series`). Pixels were assigned the agreed class C only if the consensus was reached ($N_{unique} = 1$) and supported by at least two multi-class datasets ($N_{mc} \geq 2$) (using `r.mapcalc`); otherwise, they were masked as "NoData."

Algorithm 3.1 MOLCA Generation

Require: Set of Raw Datasets $\mathcal{D}$, Harmonization Rules $\mathcal{R}$, Set of Tiles $\mathcal{T}$

Ensure: Harmonized Consensus Map (*MOLCA*)

```

1: for all Tile  $T$  in  $\mathcal{T}$  do
2:   1. Data Harmonization
3:   for all Dataset  $D \in \mathcal{D}$  intersecting  $T$  do
4:     Mosaic raw fragments of  $D$  within  $T$  (if multiple files exist)
5:     Align to  $T$  (Reprojection/Resampling)
6:     Reclassify to target legend using  $\mathcal{R}$ 
7:   end for
8:   2. Compute Cross Product and Count Map
9:   Compute cross-product of all layers:  $Map_{cross}$ 
10:  Extract unique attribute table from  $Map_{cross}$ 
11:  Calculate valid count map  $N_{mc}$ 
12:  3. Apply Rules
13:  for all Combination  $K$  in table do
14:    if Unique Class Count  $N_{unique}$  in  $K == 1$  AND  $N_{mc} \geq 2$  then
15:       $MOLCA \leftarrow$  Agreed Class  $C$ 
16:    else
17:       $MOLCA \leftarrow$  NoData
18:    end if
19:  end for
20:  4. Output
21:  Export  $MOLCA$  to GeoTIFF
22: end for

```

3.3.5. Limitations

Considering that the generation of MOLCA is based on unanimous consensus, MOLCA can be regarded as a benchmark land cover map. Nevertheless, it still has some inevitable limitations.

The unanimous consensus rule for generating MOLCA is extremely conservative, resulting in a high-confident but spatially sparse map. The result is highly affected by the accuracies

of all input datasets and sensitive to the lowest-performing input dataset; the majority was potentially omitted. Especially considering that not all the source years of the input datasets align with the target year, some pixels cannot reach agreement due to land cover change.

3.4. Probabilistic Consensus: MOLCOPA

Taking into account the aforementioned limitations of MOLCA, a probabilistic consensus rule was proposed. In spite of a relatively lower confidence, it can result in more flexible products with spatially larger coverage.

3.4.1. Overall Workflow

Figure 3.5 illustrates the overall workflow of the generation of MOLCOPA. The workflow is organized into four sequential phases:

- (1) **Target Defining:** The process began with defining the targets with referring to the targets defined for MOLCA generation previously.
- (2) **Data Preparation:** In this phase, all the data sets determined for the target years were prepared.
- (3) **Probabilistic Consensus Rule:** This was the core adjustment from MOLCA. Here, the method transited from a strict consensus-based rule to a probabilistic one, generating probability maps and a count map that capture pixel-wise class probabilities and confidence levels.
- (4) **Derive Land Cover Maps:** Starting from the probability maps and the count map, various land cover mapping strategies could be applied, which could substantially increase flexibility and scalability.
- (5) **Output and Evaluation:** The final outputs were the various land cover maps resulting from different mapping strategies. The outputs were then evaluated to ensure reliability.

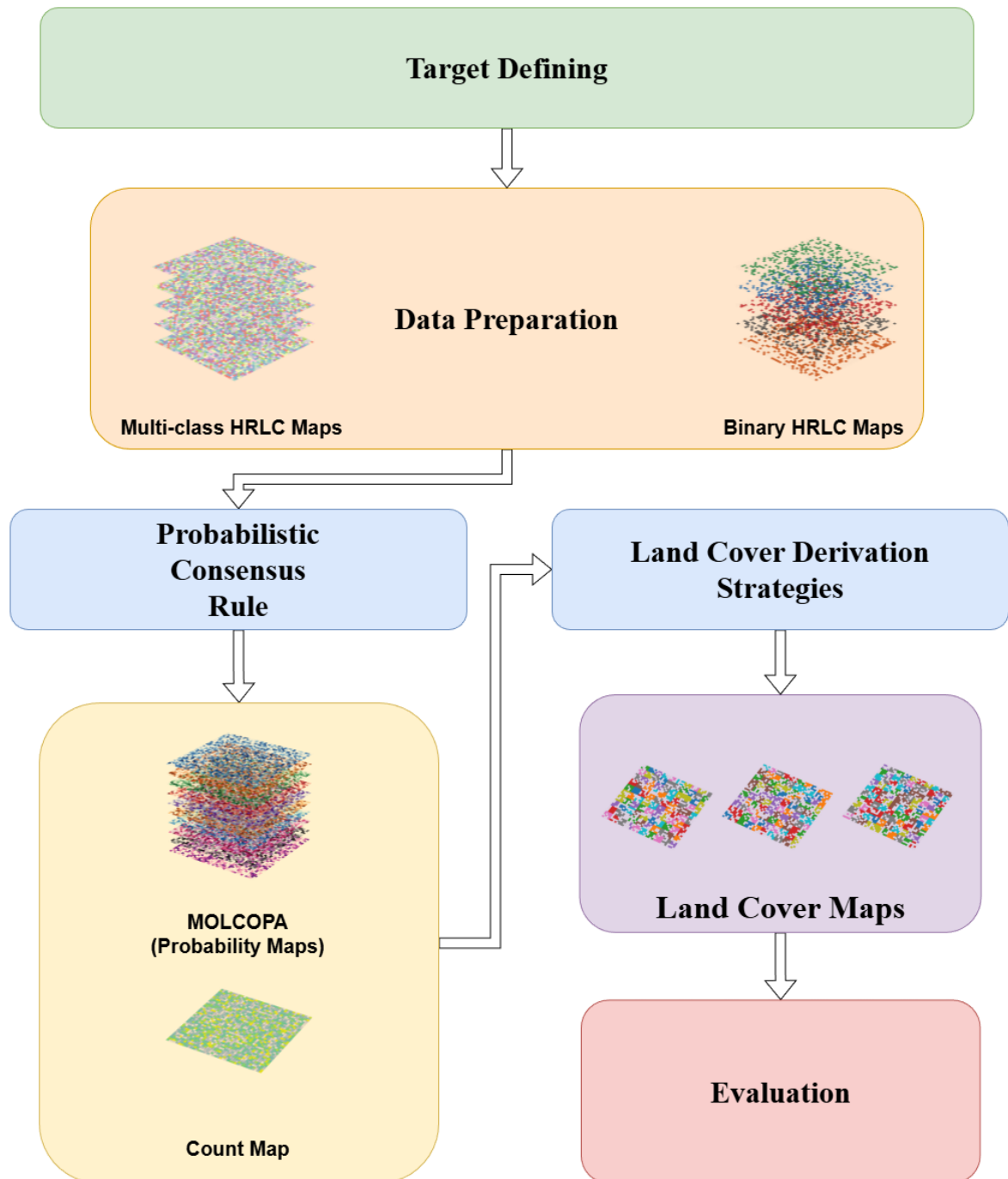


Figure 3.5: Overall Workflow of the MOLCOPA Generation

3.4.2. Target Defining and Data Preparation

Based on the quantity and variety of input datasets used to generate MOLCA, the year 2019 was selected as the only target year. There are 14 maps available for 2019 comprised of 6 multi-class HRLC products and 8 binary HRLC products. The quantity of multi-class HRLC products is vital for this approach since the probability is calculated with reference to the quantity of multi-class HRLC maps.

The data used for MOLCOPA 2019 were completely identical to the ones used for the generation of MOLCA 2019.

3.4.3. Probabilistic Consensus Rules

The generation of the MOLCOPA probability maps involved two steps: initial probability estimation derived from multi-class datasets, followed by applying logical constraints using binary datasets.

3.4.3.1. Probability Calculation

For a given pixel p , the probability of belonging to a land cover class c , which is denoted as $P(c, p)$ and calculated by Equation (3.3).

$$\begin{cases} P(c, p) = \begin{cases} \text{NoData}, & \text{if } N_{mc}(p) = 0 \\ \frac{B(c, p)}{N_{mc}(p)}, & \text{otherwise} \end{cases} \\ B(c, p) = \sum_{i=1}^{K_{mc}} \mathbb{I}(M_{mc}(i, p) = c) \\ N_{mc}(p) = \sum_{i=1}^{K_{mc}} \mathbb{I}(M_{mc}(i, p) \neq \text{NoData}) \end{cases} \quad (3.3)$$

Where:

- $P(c, p)$ is the probability of belonging to land cover class c for the pixel p .
- $B(c, p)$ is the number of multi-class maps that classify the pixel p as class c .
- $N_{mc}(p)$ is the count of valid (non-NoData) multi-class maps available at pixel p .
- $M_{mc}(i, p)$ is the class label of the i -th multi-class map at pixel p .
- K_{mc} is the number of input multi-class datasets.

- $\mathbb{I}(\cdot)$ is the indicator function.

3.4.3.2. Binary Constraints

Binary maps were used as constraints. A logical conflict occurs if the binary map asserts the presence of class c , but the multi-class consensus indicates zero probability for it. If a conflict is detected for any class c' , the final probability values for all land cover classes c at pixel p are set to "NoData."

The final probability map MOLCOPA for each class and each pixel is determined by Equation (3.4):

$$MOLCOPA(c, p) = \begin{cases} \text{NoData} & \text{if } \exists c' \in \text{Classes s.t.} \\ & \exists i \in \{1, \dots, K_{bi}\}, M_{bi}(i, p) = c' \\ & \text{AND } P(c', p) = 0 \\ P(c, p) & \text{otherwise} \end{cases} \quad (3.4)$$

Where:

- $M_{bi}(i, p)$ is the class label of the i -th binary map at pixel p .
- K_{bi} is the number of input binary datasets.

3.4.4. Derivation of Land Cover Maps

The MOLCOPA is a probabilistic stack where each pixel p contains a probability distribution for land cover classes. To generate a land cover map, a strategy must be defined.

Four different decision strategies were considered in this study. They show the trade-off between spatial coverage and confidence level.

Firstly, the maximum probability $P_{max}(p)$ and the candidate class $C_{cand}(p)$ are defined by Equation (3.5).

$$P_{max}(p) = \max_c P(c, p), \quad C_{cand}(p) = \arg \max_c P(c, p) \quad (3.5)$$

3.4.4.1. Strategy 1: Overall Agreement (MOLCA)

This strategy applies the strictest constraint, which requires unanimous consensus among all input datasets. And a pixel is classified only if all valid datasets agree on the same

label. In this case, the produced land cover map will be identical to MOLCA. The overall agreement strategy is defined by Equation (3.6).

$$LC_{OA}(p) = \begin{cases} C_{cand}(p) & \text{if } P_{max}(p) = 1.0 \quad \text{AND} \quad N_{mc}(p) \geq 2 \\ \text{NoData} & \text{otherwise} \end{cases} \quad (3.6)$$

3.4.4.2. Strategy 2: $N - 1$ Agreement

To account for potential outliers or errors in a single dataset, the $N - 1$ agreement strategy allows exactly one input map to disagree with the majority. The threshold is dynamically calculated according to N_{mc} . In addition, this strategy requires at least three multi-class datasets involved in calculation to ensure the confidence level. The $N - 1$ agreement strategy is defined by Equation (3.7):

$$LC_{N-1}(p) = \begin{cases} C_{cand}(p) & \text{if } P_{max}(p) \geq \frac{N_{mc}(p)-1}{N_{mc}(p)} \quad \text{AND} \quad N_{mc}(p) \geq 3 \\ \text{NoData} & \text{otherwise} \end{cases} \quad (3.7)$$

3.4.4.3. Strategy 3: Threshold-based Major Agreement

This strategy provides a flexible threshold-based choice. The class with highest probability that exceeds a pre-defined threshold τ (e.g., $\tau \in \{0.5, 0.6, 0.7, \dots\}$) will be assigned to the land cover map. Also, this strategy requires at least three multi-class datasets involved in calculation to ensure the confidence level. The major agreement strategy is defined by Equation (3.8):

$$LC_{major}(p) = \begin{cases} C_{cand}(p) & \text{if } P_{max}(p) > \tau \quad \text{AND} \quad N_{mc}(p) \geq 3 \\ \text{NoData} & \text{otherwise} \end{cases} \quad (3.8)$$

Thresholds τ of 0.5, 0.6, 0.7, 0.8 and 0.9 were tried in this study.

3.4.4.4. Strategy 4: Hybrid Agreement

Considering that some classes may only occur on very few multi-class datasets after legend harmonization, $N-1$ agreement and majority agreement are not appropriate enough. For example, if majority agreement strategy is applied and the threshold τ is set higher than 0.67, the land cover contains no "Shrubland" pixels. It is impossible for the probability to exceed the threshold as "Shrubland" exists only in two multi-class datasets and $N_{mc} \geq 3$

is required.

An adaptive hybrid strategy is defined to retain agreed classes as much as possible.

- (1) **In data-rich areas** ($N_{mc}(p) \geq 3$): Major Agreement strategy is applied, accommodating minor disagreement.
- (2) **In data-sparse areas** ($N_{mc}(p) = 2$): Overall Agreement strategy is applied, ensuring reliability.

The hybrid strategy is defined by Equation (3.9):

$$LC_{hybrid}(p) = \begin{cases} C_{cand}(p) & \text{if } (N_{mc}(p) \geq 3 \text{ AND } P_{max}(p) > \tau) \\ & \text{OR } (N_{mc}(p) = 2 \text{ AND } P_{max}(p) = 1.0) \\ \text{NoData} & \text{otherwise} \end{cases} \quad (3.9)$$

The Threshold $\tau = 0.67$ was determined based on the results of Threshold-based Major Agreement.

3.4.5. Implementation

The proposed MOLCOPA generation workflow was implemented using GRASS GIS driven by Python scripting. The generation of MOLCOPA follows Algorithm 3.2. The processing was carried out tile-by-tile with the tiling system mentioned above.

Firstly, data harmonization was performed in the same way applied in MOLCA.

Next, Probability calculation was conducted using the harmonized multi-class layers. The count map N_{mc} was generated using `r.series`. Then the class occurrence for each pixel $N_{mc}(c, p)$ was calculated to determine the probability $P(c, p)$ using `r.mapcalc`.

Subsequently, binary constraints were applied by using `r.mapcalc`. It is worth noting that a binary dataset can be used multiple times with different c_{target} .

Finally, various land cover maps were derived from the probability maps and count map using `r.mapcalc`. The probability maps, the count map and the land cover maps were then exported.

Algorithm 3.2 MOLCOPA Generation

Require: Multi-class Datasets $\mathcal{M}_{mc}$, Binary Constraint Maps $\mathcal{M}_{bi}$, Harmonization Rules $\mathcal{R}$, Set of Tiles $\mathcal{T}$

Ensure: Probability Maps *MOLCOPA*, Count Map N_{mc} and Consensus Map *LC*

```

1: for all Tile  $T$  in  $\mathcal{T}$  do
2:   1. Data Harmonization
3:   for all Dataset  $D \in (\mathcal{M}_{mc} \cup \mathcal{M}_{bi})$  intersecting  $T$  do
4:     Mosaic raw fragments of  $D$  within  $T$  (if multiple files exist)
5:     Align to  $T$  geometry (Reprojection/Resampling)
6:     Reclassify using  $\mathcal{R}$  to target legend  $\rightarrow$  Layer  $L_{mc}$  or  $L_{bi}$ 
7:     Register  $L_D$  as Multi-class type or Binary type
8:   end for
9:   2. Probability Calculation
10:  Calculate Valid Count:  $N_{mc}(p)$ 
11:  for all Class  $c$  in Legend do
12:    Count class occurrences:  $B(c, p)$ 
13:    Compute Probability:  $P(c, p) \leftarrow B(c, p)/N_{mc}(p)$ 
14:  end for
15:  3. Apply Binary Constraints
16:  for all Binary Map  $L_{bi}$  targeting Class  $c_{target}$  do
17:    if  $L_{bi}(p) == c_{target}$  AND  $(P(c_{target}, p) == 0$  OR NoData) then
18:      Conflict Detected: Binary confirms class, but Multi-class consensus denies
        it.
19:      Set probabilities of all classes at pixel  $p$  to NoData.
20:    end if
21:  end for
22:  4. Deriving LC Maps
23:  Identify max probability:  $P_{max}(p) = \max_c P(c, p)$ 
24:  Identify candidate class:  $C_{cand}(p) = \arg \max_c P(c, p)$ 
25:  Apply decision rules (e.g., Overall Agreement, N-1 Threshold) to derive LC
26:  Output Export MOLCOPA , Count Map  $N_{mc}$  and Land Cover Maps LC to
    GeoTIFF.
27: end for

```

3.5. GeoAI-based Land Cover Mapping

After generating the aforementioned products, benchmark land cover maps with different confidence levels were obtained.

Thus, it could be utilized to extract training samples for GeoAI-based land cover mapping, avoiding manually collecting training samples. Especially with released embedding data, some GeoAI-based methods are becoming easier to implement. Geo-AI-based land cover mapping is able to produce an almost complete land cover map, which means the coverage on the study area would be close to 100%.

In this study, three main approaches were considered: unsupervised, semi-supervised machine learning method and deep learning method. Since MOLCOPA was only generated for the year 2019, the GeoAI-based methods aimed at the year 2019 as well.

3.5.1. Unsupervised Approach

3.5.1.1. Overall Workflow

Figure 3.6 shows the overall workflow of the unsupervised LC mapping approach. It can be subdivided into 4 phases:

- (1) **Feature Selection:** The process began with selecting proper features for the clustering algorithm.
- (2) **Clustering:** In this phase, a clustering algorithm was applied with the selected feature.
- (3) **Automatic Labeling:** The result of clustering algorithm was several anonymous clusters. Thus, a strategy should be applied to have the cluster labeled.
- (4) **Output and Evaluation:** The final output was a complete land cover map. The outputs were then evaluated to ensure reliability.

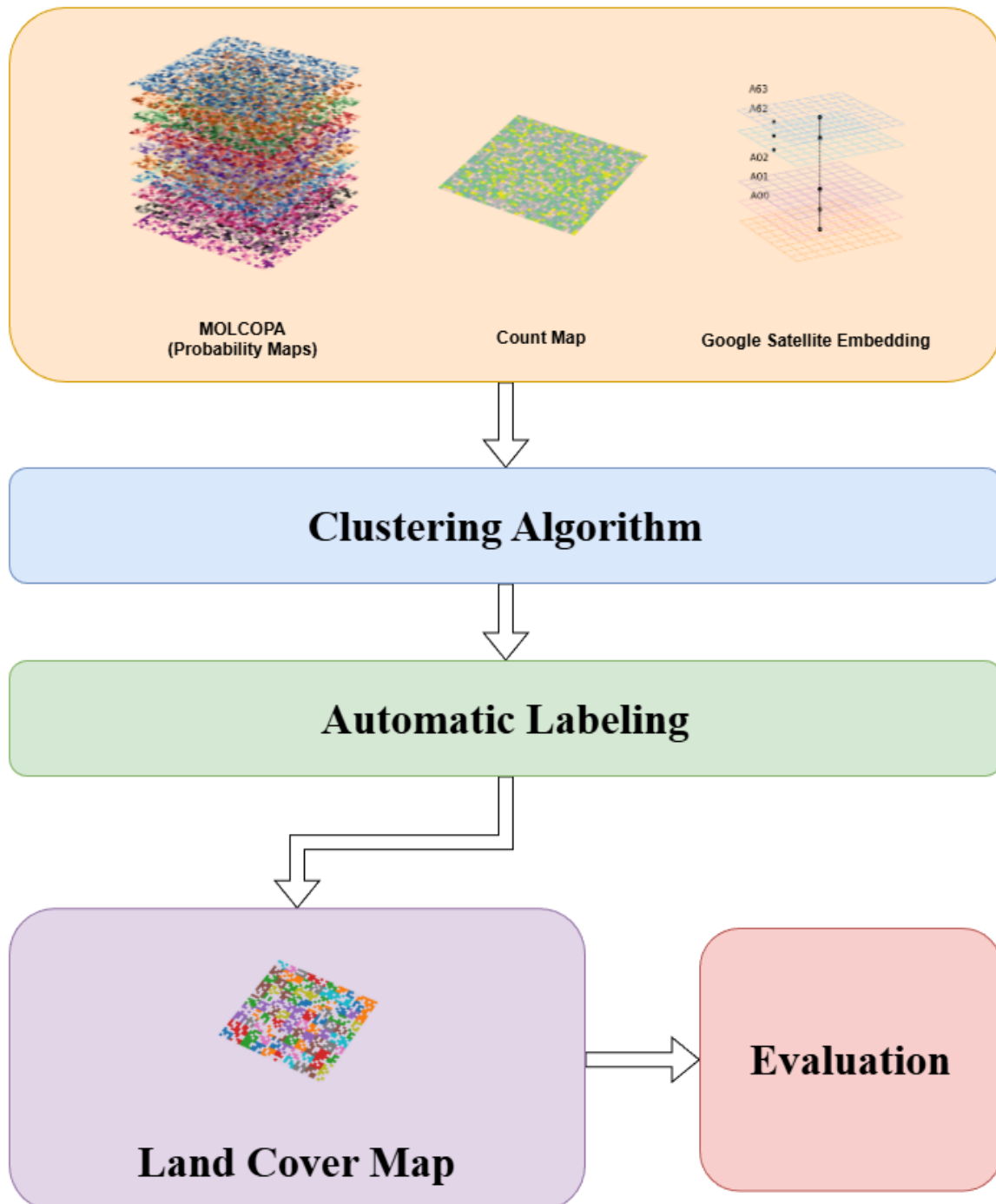


Figure 3.6: Overall Workflow of Unsupervised Approach

3.5.1.2. Feature Selection and Clustering

Instead of using satellite images as features, foundation model embedding datasets were considered, which has been proven to be effective in a land cover mapping task [21]. One of the most important advantages of using embedding datasets is that it has outstanding performance with sparse label data [7, 21].

There are three mainstream foundation model embedding datasets: Google Satellite Embedding [7], Tessera Embedding [21] and Presto Embedding [46]. Google Satellite Embedding is the only one that provides the coverage of the study area for the target year 2019 (as of January 2026), thus it was selected to be used.

MOLCOPA and the Count Map N_{mc} were also added as features to provide additional information that shows the probability and confidence level.

K-Means was selected to be used in this study considering the low computational complexity. The effectiveness of K-means in the field of land cover has been proven in some previous studies [31, 32].

3.5.1.3. Automatic Labeling

Figure 3.7 illustrates the process of the auto labeling strategy in this study. After executing the clustering algorithm, a map with anonymous clusters can be obtained. Thus, labeling the clusters is the main challenge. The auto-labeling was achieved through calculating mean probability based on MOLCOPA. For each anonymous cluster, a mean probability vector was built, containing the likelihood of belonging to each class. Two distinct computation strategies are defined for the mean vector:

- (1) **Global Auto-Labeling** : In the global approach, the mean probability vector for each cluster is calculated using all pixels within the entire study area that belong to that cluster.
- (2) **Local Auto-Labeling** : In the Local approach, the study area is divided into smaller spatial windows (e.g. $10km \times 10km$ squares), and the auto-labeling logic is applied independently within each tile.

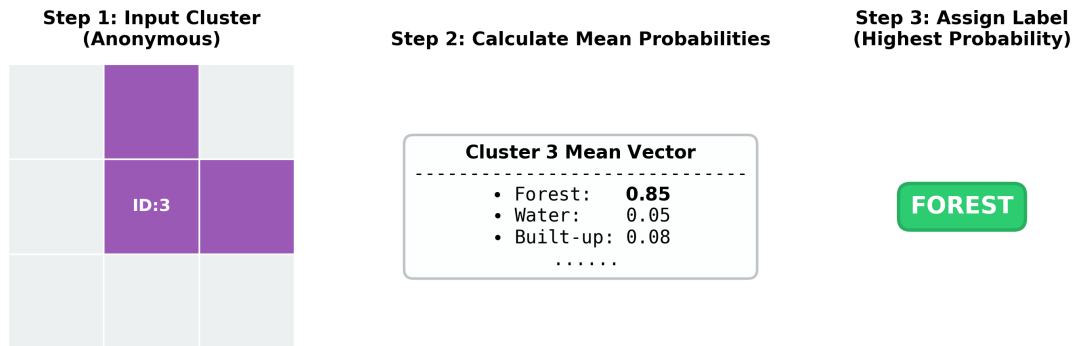


Figure 3.7: Auto Labeling Strategy

3.5.1.4. Experiment Design and Implementation

The implementation relied on the Google Earth Engine Python API and Colab.

"NoData" pixels caused by the binary constraints in the MOLCOPA probability bands were filled with an equal probability of 10% across all 10 bands. L2-normalization was applied to the feature vectors, making Euclidean distance equivalent to Cosine similarity for K-means clustering.

K-means was implemented using `ee.Clusterer.wekaKMeans` with 10 clusters. Global and local auto-labeling were both used and compared. For local auto-labeling, the spatial window sizes of 5×5 km and 10×10 km were tried. Auto-Labeling was performed using Python with Rasterio in the local environment (Windows PC). "NoData" pixels in the MOLCOPA were not filled for auto-labeling because it is meaningless.

3.5.2. Semi-supervised Approach

3.5.2.1. Overall Workflow

Figure 3.8 shows the overall workflow of the semi-supervised LC mapping approach. It can be subdivided into 4 phases:

- (1) **Sample Extraction:** The process began with extracting samples from MOLCA.
- (2) **Feature Selection:** In this phase, a suitable feature vector was selected.
- (3) **Supervised Classification:** Different classifiers were trained with the extracted samples from MOLCA and the feature vector.
- (4) **Output and Evaluation:** The trained classifier were used to make a global pre-

diction for the study area, thus a complete land cover map is obtained. The output was then evaluated to ensure reliability.

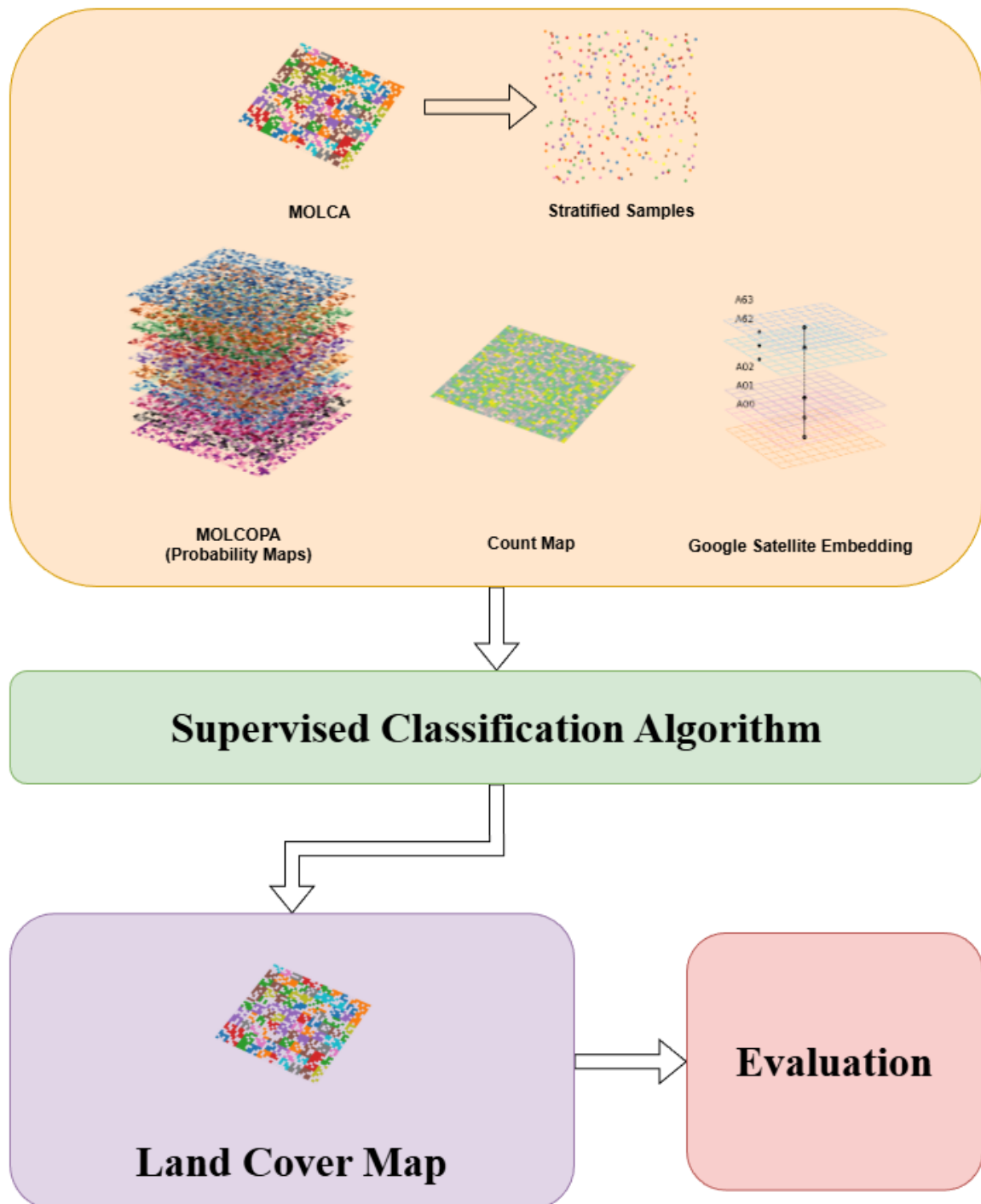


Figure 3.8: Overall Workflow of Semi-supervised Approach

3.5.2.2. Feature Selection and Sample Extraction

Google Satellite Embedding, MOLCOPA, and the count map were selected as potential features that can be applied in this semi-supervised approach, considering the characteristics mentioned in Section 3.5.1.

Given that the MOLCA product demonstrates sufficient reliability to serve as a benchmark land cover map, it was utilized as the source for extracting the training samples. Specifically, training samples were extracted directly from the MOLCA using a stratified sampling approach. Class sizes were balanced to mitigate potential issues caused by class imbalance during model training unless a specific class had insufficient samples, in which case its sample size was set to the maximum available count. Since these sample points were not gold standard data that can fully represent the ground truth, they were named "silver points" in this study.

Three distinct training datasets were constructed by performing stratified sampling on the MOLCA 2019, with target sample sizes of 50, 100, and 200 pixels per class, respectively. These datasets are denoted as Set-50, Set-100, and Set-200. As only 7 valid pixels were available in the MOLCA, all 7 "Shrubland" pixels were included in every dataset to maximize representation. In addition, since no pixel was classified as "Permanent ice and snow" and "Lichens and Mosses" in MOLCA, no samples of these two classes were extracted.

Figures 3.9, 3.10, and 3.11 show the spatial distribution and the class count of Set-50, Set-100, and Set-200 respectively.

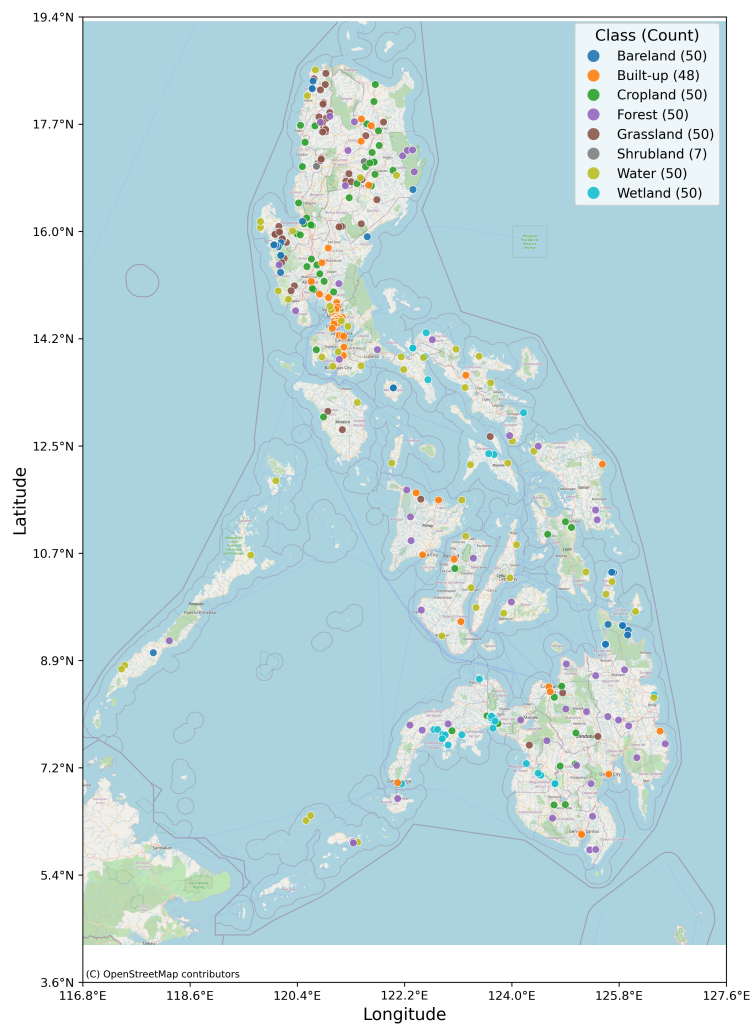


Figure 3.9: Spatial Distribution and Class Count of Set-50

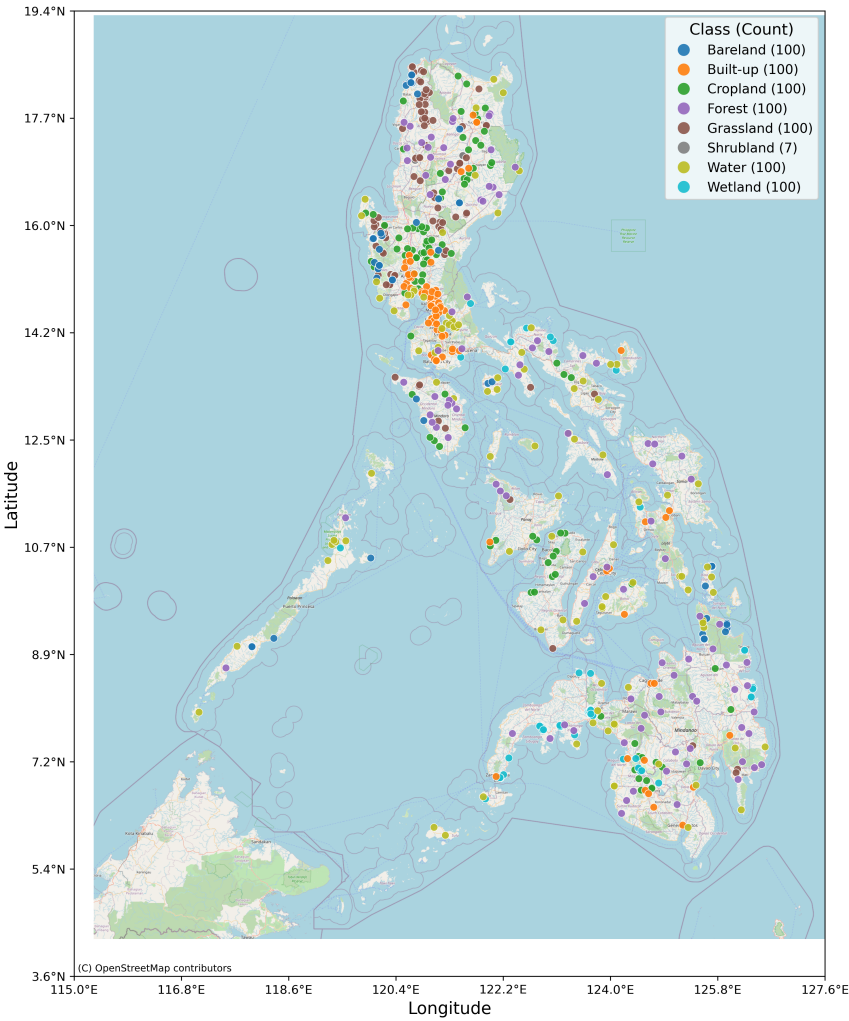


Figure 3.10: Spatial Distribution and Class Count of Set-100

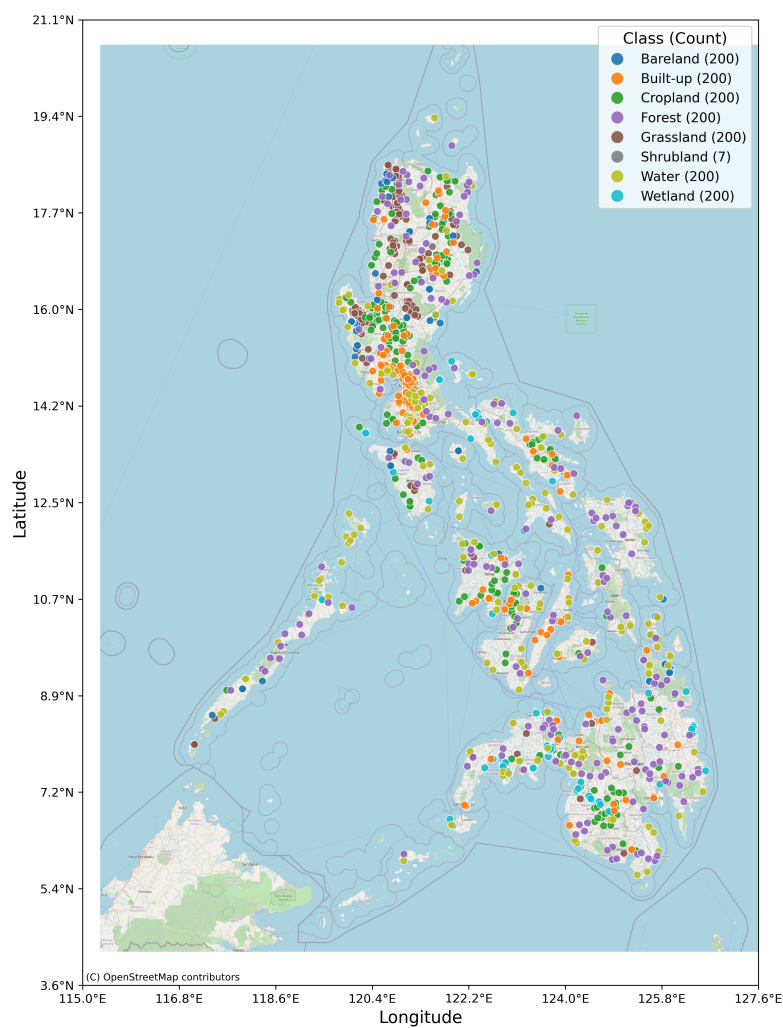


Figure 3.11: Spatial Distribution and Class Count of Set-200

3.5.2.3. Classifier Selection

To determine the optimal model, different classifiers were evaluated.

Since RF and SVM are the most commonly used classic classifiers for the land cover mapping of remote sensing data [47], they were tried in this study. MLP was also tried, as the effectiveness of applying MLP with embedding data has been verified [21].

3.5.2.4. Experiment Design

A multifactorial preliminary experiment was designed to explore the optimal configurations for the semi-supervised approach. Table 3.2 illustrates the overall configuration of the experiments.

The experiment includes the following three variables:

- (1) **Feature Configuration:** To assess the efficacy of MOLCOPA and the count map, two configurations were compared:
 1. Embeddings Only: Using strictly Google Satellite Embeddings to enforce semantic learning.
 2. Full Stack: Combining Embeddings, MOLCOPA, and the Count map to test for feature dominance versus complementary enhancement.
- (2) **Classifiers:** Three distinct classifiers were evaluated: Random Forest, SVM, and MLP.
- (3) **Sample Size:** Considering that different algorithms exhibit varying sensitivities to the size of the training dataset [47], three stratified datasets were constructed: Set-50, Set-100 and Set-200 as mentioned above.

The overall performance of all experimental groups was evaluated. The results were summarized to identify the optimal configuration for each classifier that maximizes accuracy.

Group ID	Feature Configuration	Classifier	Training Dataset
RF-50A	Embeddings Only	Random Forest	Set-50
RF-100A			Set-100
RF-200A			Set-200
SVM-50A		SVM	Set-50
SVM-100A			Set-100
SVM-200A			Set-200
MLP-50A		MLP	Set-50
MLP-100A			Set-100
MLP-200A			Set-200
RF-50B	Full Stack	Random Forest	Set-50
RF-100B			Set-100
RF-200B			Set-200
SVM-50B		SVM	Set-50
SVM-100B			Set-100
SVM-200B			Set-200
MLP-50B		MLP	Set-50
MLP-100B			Set-100
MLP-200B			Set-200

Table 3.2: Summary of Experimental Configurations

3.5.2.5. Implementation

The implementation relied on the Google Earth Engine Python API, Colab, and some relevant Python libraries (geemap, scikit-learn, imbalanced-learn, and TensorFlow).

Because some pixels in MOLCOPA were set to "NoData" when the binary constraint was applied, preprocessing was required to produce a complete land cover map. "NoData" in the MOLCOPA probability bands was filled with an equal probability of 10% across all 10 bands.

For RF and SVM, Synthetic Minority Over-sampling Technique (SMOTE) [9] (using `imblearn.over_sampling.SMOTE`) was applied to oversample the training dataset to address the issue of imbalanced training datasets (only 7 "Shrubland" samples). This process generated synthetic samples for minority classes, resulting in a fully class-balanced training set. For all classifiers, features were standardized before training.

SVM was implemented using `ee.Classifier.libsvm` with RBF kernel.

RF was implemented using `ee.Classifier.smileRandomForest` with the number of trees in the forest set to 1000.

MLP was implemented using TensorFlow. As shown in Figure 3.12, the MLP model features two fully connected hidden layers with ReLU activation and Dropout regularization to prevent overfitting, ending with a Softmax output layer. The model was compiled using the Adam optimizer with "sparse categorical cross-entropy" as the loss function and accuracy as the evaluation metric. The training process consisted of 50 epochs with a batch size of 32. A validation split of 10% was applied to the training set to monitor model performance during training.

The global prediction was performed with three classifiers with their optimal configurations, producing three land cover maps. Then further analysis was conducted on these three maps.

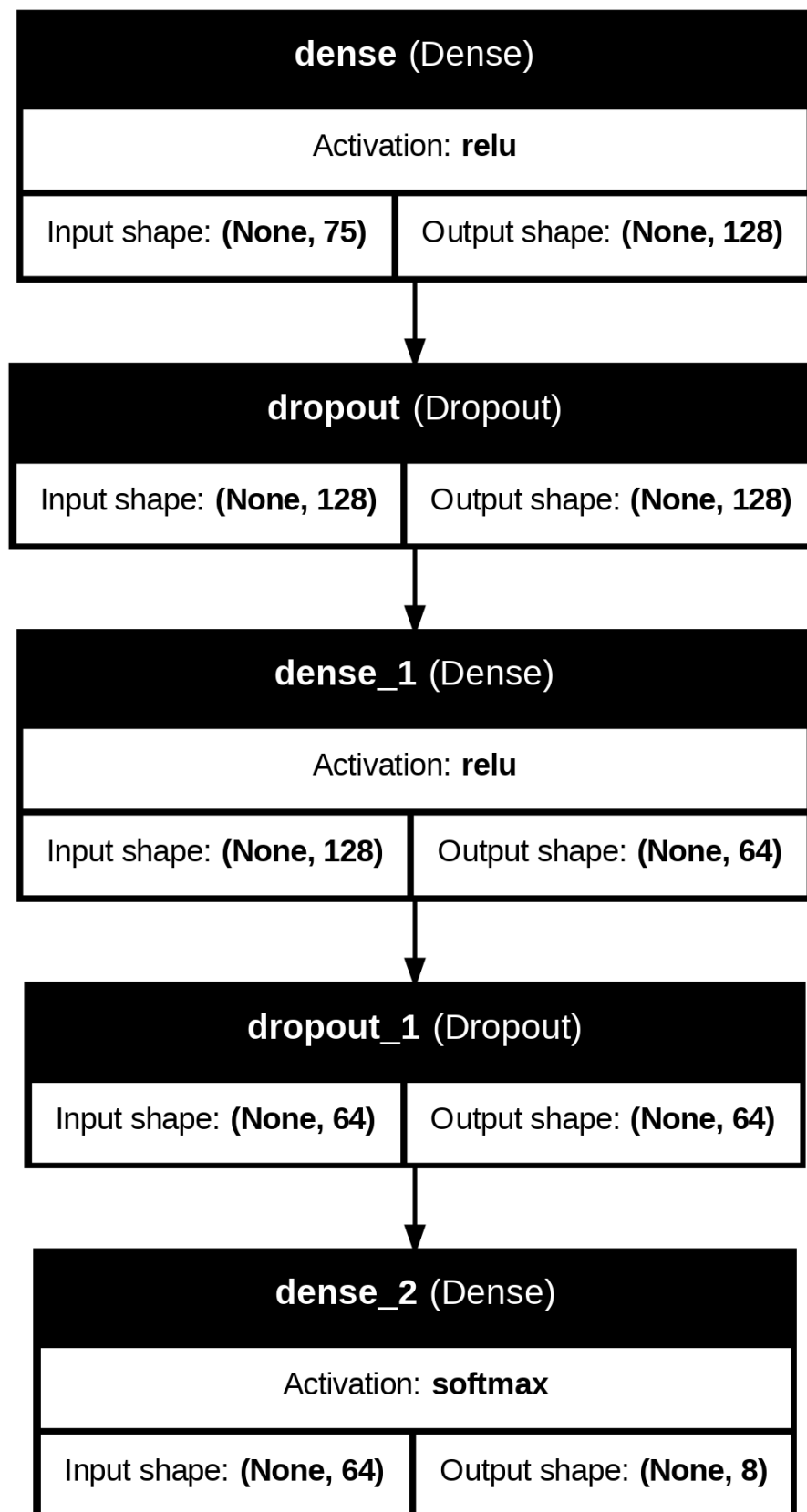


Figure 3.12: MLP Model

3.6. Evaluation

In this section, the evaluation methods for above-mentioned land cover maps are included.

3.6.1. MOLCA and MOLCOPA

To evaluate the MOLCA, both statistical analysis and accuracy assessment were performed to evaluate the trade-off between coverage and accuracy.

For the statistical analysis, all tiles of the land cover map to be evaluated were merged and then the pixel count for each class was calculated to detect potential anomalies.

The accuracy assessment was conducted through visual interpretation utilizing Open Foris Collect [37] and Collect Earth [2]. Open Foris Collect was used to design the survey for Collect Earth. The samples were imported to Collect Earth to perform the visual interpretation, in which Google Earth image, NDVI from MODIS, LANDSAT 7/8, and Sentinel-2 were provided.

The sample points were extracted from the generated land cover maps using the Rasterio library in Python. To ensure statistical validity, the total sample size was determined using Cochran's formula. Based on a confidence level of 95% ($Z = 1.96$), a margin of error of 0.05, and a conservative population proportion estimate of 0.5, 384 sample points were required for each map. A stratified sampling strategy was employed to address class imbalances. The initial target was an equal allocation of samples across all land cover classes. However, for classes with insufficient pixels to meet the requirement, the maximum number of available pixels was used, and the remaining sample slots were redistributed among the other classes.

The external validation dataset SRS_val was also applied for the year 2019 to improve reliability. The external dataset was clipped to the boundary of the study area and reclassified according to the harmonization rules. The points from the maps were sampled according to the validation points' coordinates. If the sample points from HRLC represents "NoData," the points would be considered invalid. Then invalid sample points were removed.

MOLCA 2000 was not evaluated due to the lack of high-resolution images on Google Earth; it was assumed to have a comparable accuracy to other MOLCAs with 30-m resolution.

The evaluation of MOLCOPA-derived land cover maps followed the same procedure as that of the MOLCA. The validation against SRS_val was also conducted to make a

comparison.

Table 3.3 presents the class distribution of the visual interpretation samples used to evaluate MOLCA and MOLCOPA-derived LC products. The corresponding spatial distributions are provided in Appendix A (Figures A.1– A.10).

Confusion matrices were created to calculate the metrics.

	Shrubland	Grassland	Cropland	Wetland	Bareland	Built-up	Water	Forest
MOLCA 2010	2	54	55	55	54	54	55	55
MOLCA 2015	48	48	48	48	48	48	48	48
MOLCA 2019	7	54	54	54	53	54	54	54
Hybrid	7	54	54	54	53	54	54	54
N-1	48	48	48	48	48	48	48	48
Major($\tau = 0.5$)	48	48	48	48	48	48	48	48
Major($\tau = 0.6$)	48	48	48	48	48	48	48	48
Major($\tau = 0.7$)	0	55	55	55	54	55	55	55
Major($\tau = 0.8$)	0	55	55	54	55	55	55	55
Major($\tau = 0.9$)	0	55	55	54	55	55	55	55

Table 3.3: Samples' Distribution for MOLCA and MOLCOPA Evaluation

3.6.2. GeoAI-based Land Cover Maps

Based on the previous accuracy assessment conducted for land cover maps derived from MOLCOPA, adequate samples had been collected. Thus, a reliable validation set for 2019 was established with them, referencing a previous study [4].

The study area was divided into "easy-to-classify" (MOLCA2019 non-NoData) and "hard-to-classify" (MOLCA2019 NoData) areas, and 384 non-repetitive sample points from each area were selected. The number of samples was also determined by Cochran's formula to make separate analysis for 2 areas available. In order to achieve a balanced distribution among land cover classes, each class had the same number of samples; the maximum available number was used and the remaining sample slots were redistributed among other classes when samples of a specific class were insufficient. Additionally, the sample points existing in the training sets (Set-50, Set-100, Set-200) were not included in this validation set. This validation set with 768 samples was denoted ValiSet-O.

The spatial distribution and the class count distribution is shown in Figure 3.13. Since no pixel was classified as "Permanent ice and snow" and "Lichens and Mosses" in the land cover maps, no samples of these two classes were extracted.

Limitations of ValiSet-O: Sample points located in the "easy-to-classify" area may still be difficult to classify, as MOLCA is not absolutely accurate; consequently, the true labels

for these points may have been re-labeled during the visual interpretation process. This is particularly evident for the "Shrubland" class, where all samples within the "easy-to-classify" subset were re-labeled. This is because MOLCA only had 7 pixels of this class, all of which were used in the training sets and therefore excluded from ValiSet-O. Conversely, sample points in the "hard-to-classify" area might only show a difference in one HRLC dataset and may not be that difficult to classify after all. In general, ValiSet-O is a very robust validation set.

ValiSet-O and SRS_val were used to assess the accuracy of the land cover maps produced. Confusion matrices were created to calculate the metrics.

A comparison between the optimal model and existing multi-class HRLC datasets was conducted using ValiSet-O. The processing workflow follows the same methodology used for validating existing datasets with SRS_val, as described in Section 2.7.

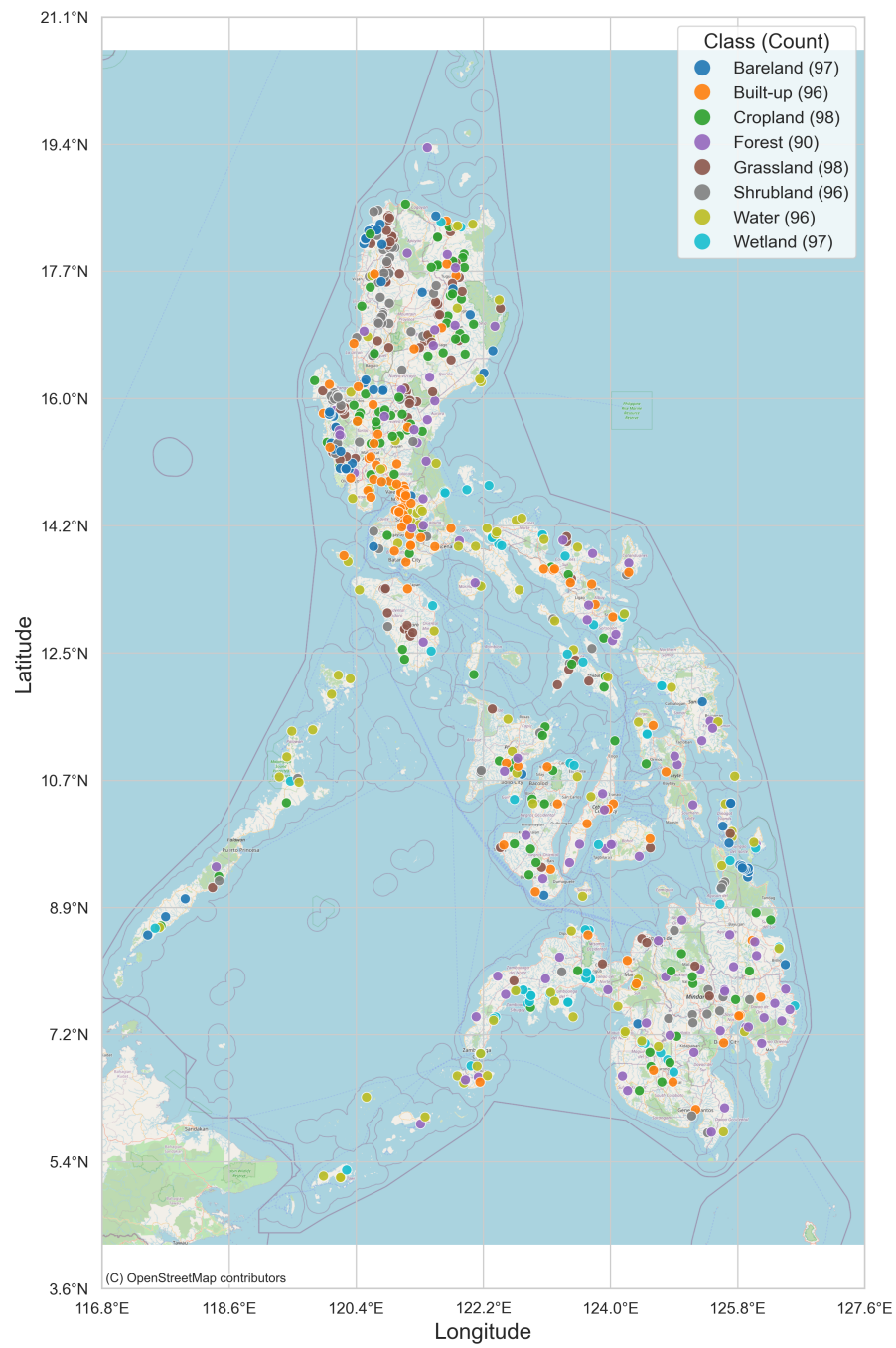


Figure 3.13: Spatial Distribution and Class Count of ValiSet-O

4 | Results

4.1. MOLCA

This section includes the analyses on the MOLCA 2000, 2010, 2015, and 2019.

4.1.1. Maps

The MOLCAs for the target years 2000, 2010, 2015, 2019 are shown in Figures 4.1, 4.2, 4.3 and 4.4.

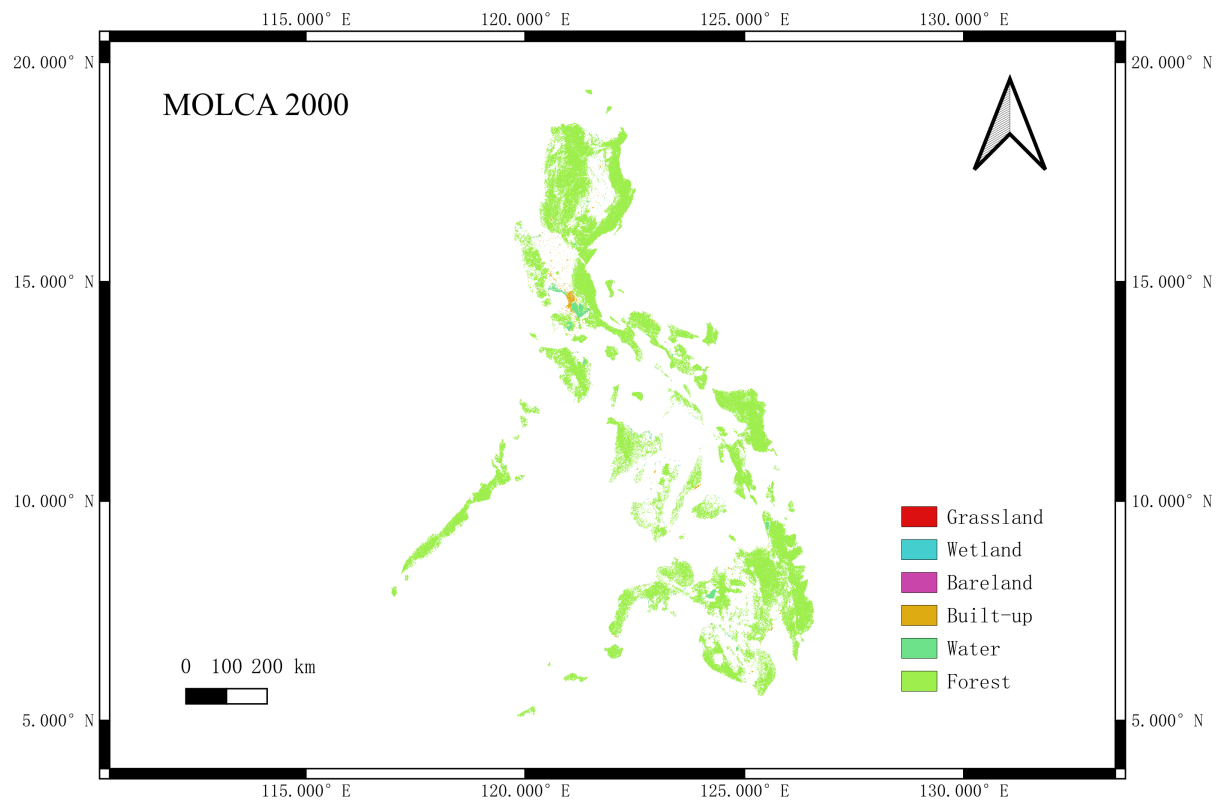


Figure 4.1: MOLCA 2000

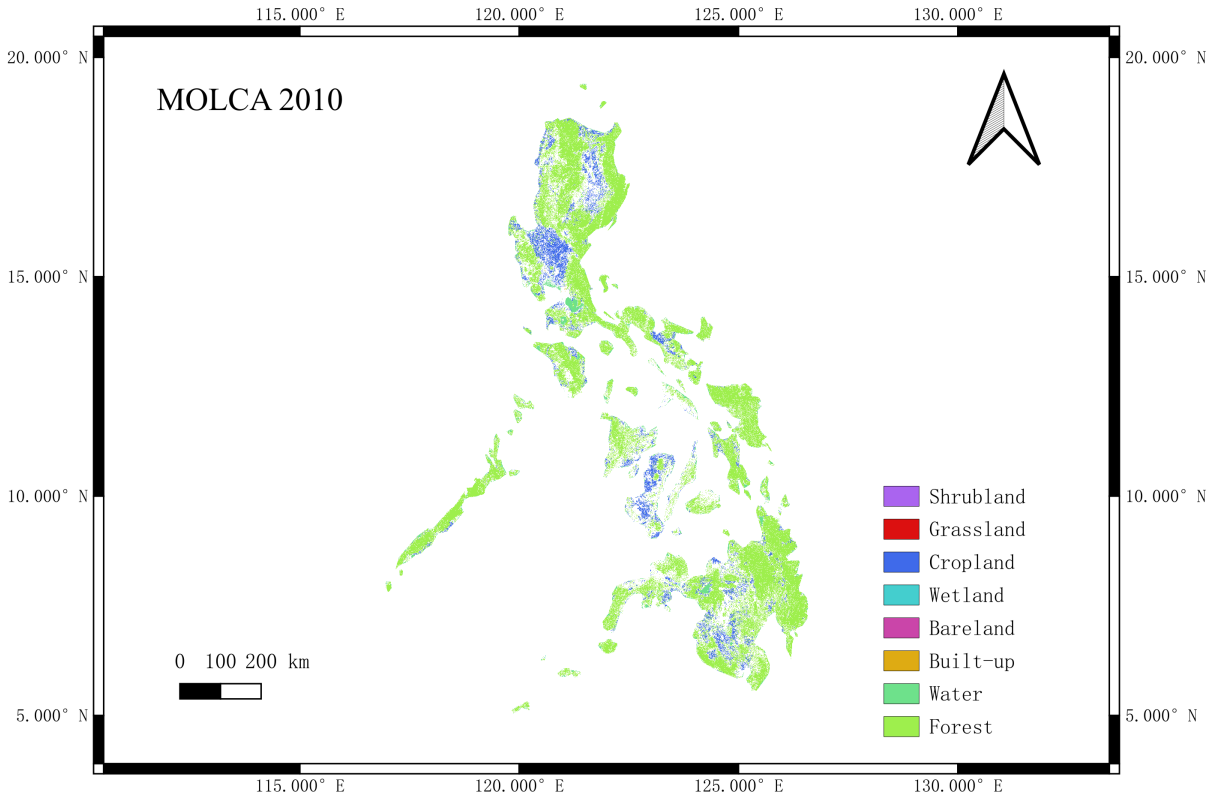


Figure 4.2: MOLCA 2010

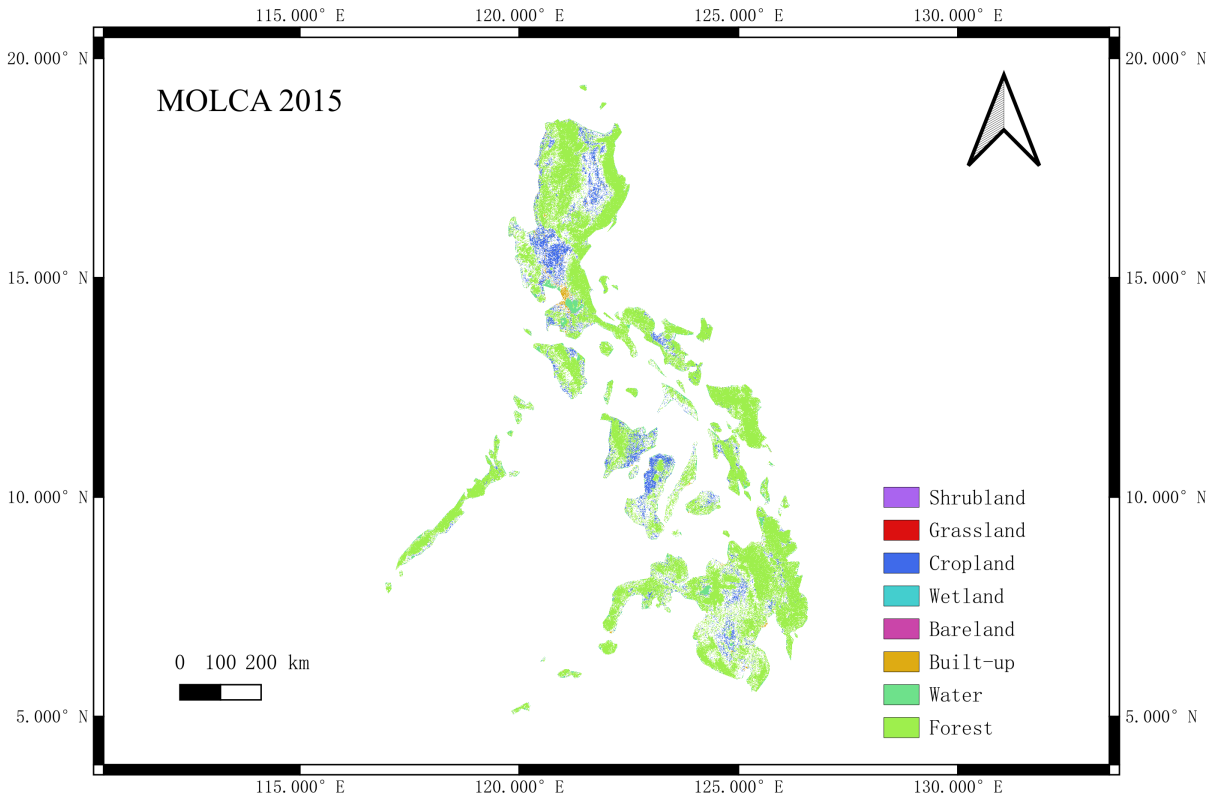


Figure 4.3: MOLCA 2015

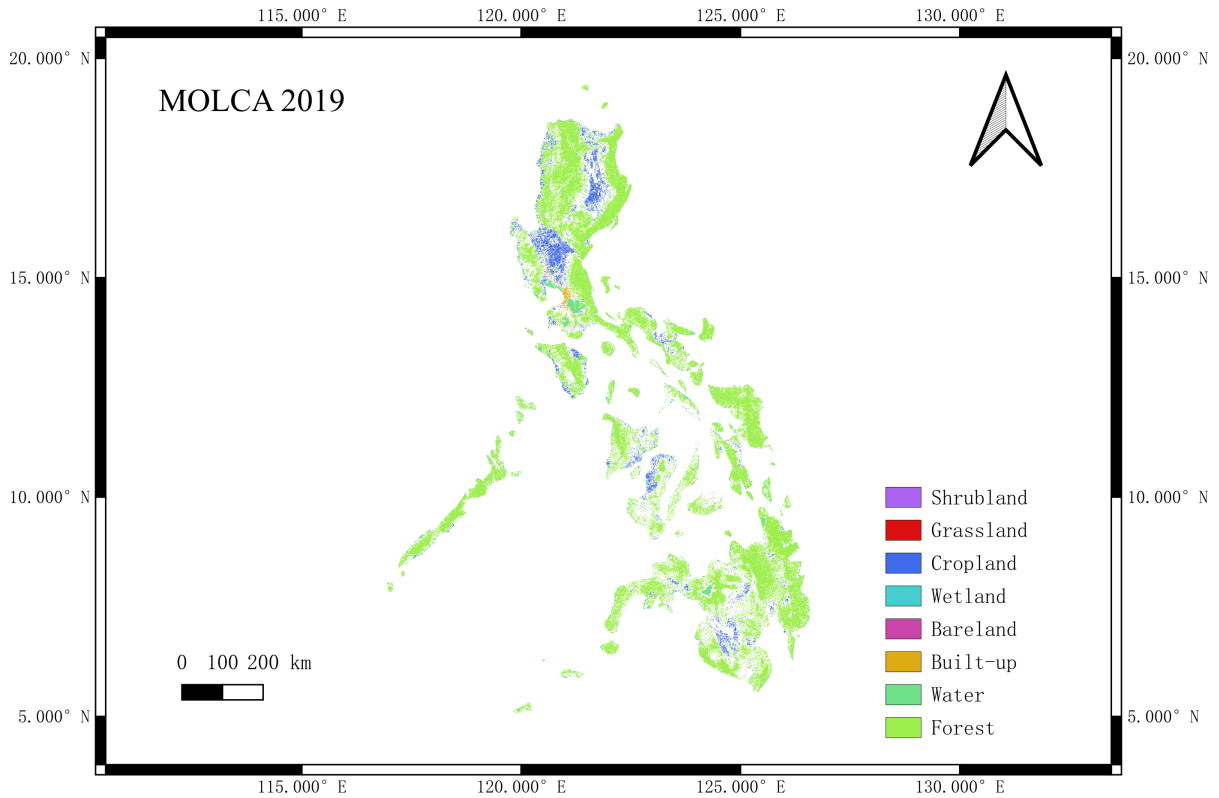


Figure 4.4: MOLCA 2019

4.1.2. Statistical Analysis

Tables 4.1 and 4.2 show the pixel count and percentage statistics for each class for MOLCA 2000, 2010, 2015, and 2019.

MOLCA for four target years shows a similar coverage level. A significant portion of the study area (ranging from 32.24% to 41.81%) is classified as "NoData". This category represents areas where input land cover maps could not reach a consensus. The high percentage in 2019 (41.81%, approximately 1.2 billion pixels) primarily results from the significantly higher number of input datasets compared to other target years, which increases the difficulty of reaching a unanimous consensus. The year 2000 (41.21%) indicates a substantial degree of disagreement, mainly because of the class "Cropland". In the legend harmonization rules, only one dataset has class "Cropland" (FROM-GLC), thus it is impossible to reach a consensus on "Cropland."

"Forest" and "Cropland" are the dominant land cover classes, though their proportions fluctuate. Classes such as "Grassland", "Bareland", "Wetland", and "Shrubland" consistently occupy very small proportions of the study area (mostly <0.1%), due to the fact that they are the minor land cover classes in this region. "Lichens and Mosses" and

"Permanent ice and snow" are missing due to the fact that they do not exist in the study area.

Different combinations of input HRLC datasets and the quality of input datasets also highly influence the results. Although the input datasets for 2010 and 2015 differ by only a single data source, significant discrepancies can be observed in the results for minority classes. Specifically, the pixel counts for "Grassland" and "Bareland" in 2010 were approximately half of those in 2015, whereas the "Wetland" exhibited a significantly higher pixel count in 2010 compared to 2015. There is almost no agreement for "Shrubland" in 2010, which sharply contrasts with the 38,832 pixels in 2015. The "Built-up" area in 2010 is significantly lower than in other years. This is primarily attributed to a misclassification of Metro Manila in the FROM-GLC 2010 dataset, where it was classified as "silt," which corresponds to "bareland" in the MOLCA legend.

The transition from 30-m (2000-2015) to 10-m (2019) resolution results in a roughly 9-fold increase in the total pixel count (from ~ 322 million to ~ 2.9 billion). Although the general composition of land cover remains similar, the 2019 map provides a more granular distribution, which is reflected in the proportions of the dominant classes.

Class	2000	2010	2015	2019
Forest	183,650,101	167,375,075	179,923,928	1,423,007,516
No Data	132,856,146	119,216,941	103,952,722	1,215,388,342
Cropland	0	26,124,761	25,048,414	162,146,777
Water	3,479,962	9,398,965	10,762,492	91,150,155
Built-up	2,140,791	89,725	2,595,009	15,470,740
Grassland	51	2,344	4,545	26,738
Bareland	243	1,756	4,428	4,283
Wetland	277,596	195,321	74,520	3,732
Shrubland	0	2	38,832	7
Total	322,404,890	322,404,890	322,404,890	2,907,198,290

Table 4.1: Pixel Count Statistics for each Land Cover Class for MOLCA

Class	2000	2010	2015	2019
Forest	56.96%	51.91%	55.81%	48.95%
No Data	41.21%	36.98%	32.24%	41.81%
Cropland	0.00%	8.10%	7.77%	5.58%
Water	1.08%	2.92%	3.34%	3.14%
Built-up	0.66%	0.03%	0.80%	0.53%
Grassland	<0.01%	<0.01%	<0.01%	<0.01%
Bareland	<0.01%	<0.01%	<0.01%	<0.01%
Wetland	0.09%	0.06%	0.02%	<0.01%
Shrubland	0.00%	<0.01%	0.01%	<0.01%

Table 4.2: Percentage Statistics for each Land Cover Class for MOLCA

4.1.3. Accuracy Assessment

The confusion matrices, Overall Accuracies, and Kappa Coefficients for MOLCA 2010, 2015, and 2019 are shown in Figures 4.5, 4.6, and 4.7, respectively. The confusion matrices for MOLCA 2010 and 2015 are derived from the visual interpretation of sampled points. For MOLCA 2019, two confusion matrices were produced: one based on visual interpretation and another based on validation against the SRS_val dataset.

In general, MOLCA for every target year shows good performance. The overall accuracy of MOLCA 2019 is the best (92.71% by visual interpretation and 94.34% on SRS_val), mainly attributed to the highest number of input datasets (6 multi-class and 8 binary HRLC datasets); particularly, it has a better spatial resolution of 10-m compared to others. The MOLCA 2010 has the worst performance from the perspective of overall accuracy, probably resulting from the poor quality of HRLC datasets for early years and a lack of HRLC datasets. Overall, the overall accuracy is strongly correlated with the number of input multi-class HRLC datasets.

In MOLCA 2019, nearly all classes are correctly classified, with the exception of "Shrubland" and "Grassland." In contrast, MOLCA 2010 and 2015 exhibit varying degrees of misclassification across many categories, some of which were counter-intuitive. This can be related to the relatively coarser 30-m resolution. Within a 30 m $\times$ 30 m area, multiple land cover classes often coexist, which tends to cause different classification results.

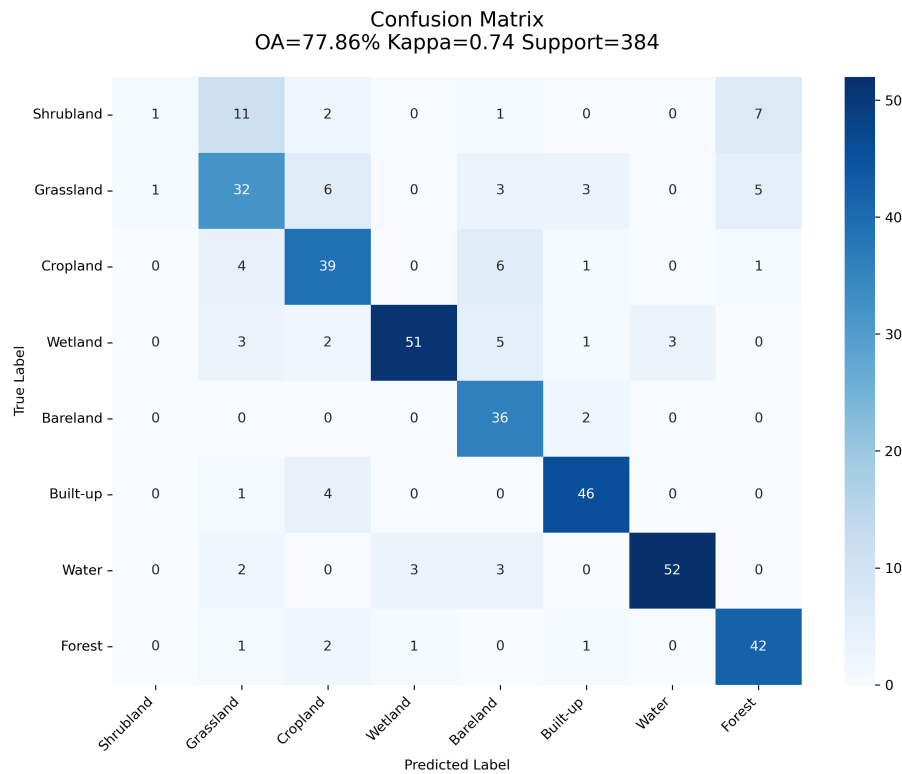


Figure 4.5: Confusion Matrix for MOLCA 2010 (Interpretation)

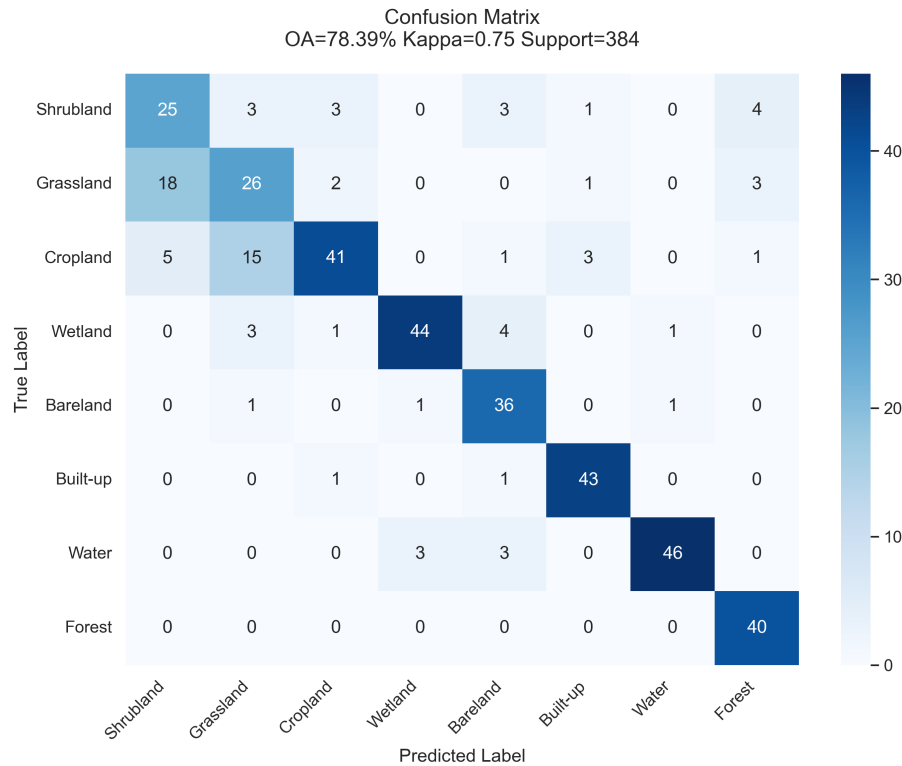


Figure 4.6: Confusion Matrix for MOLCA 2015 (Interpretation)

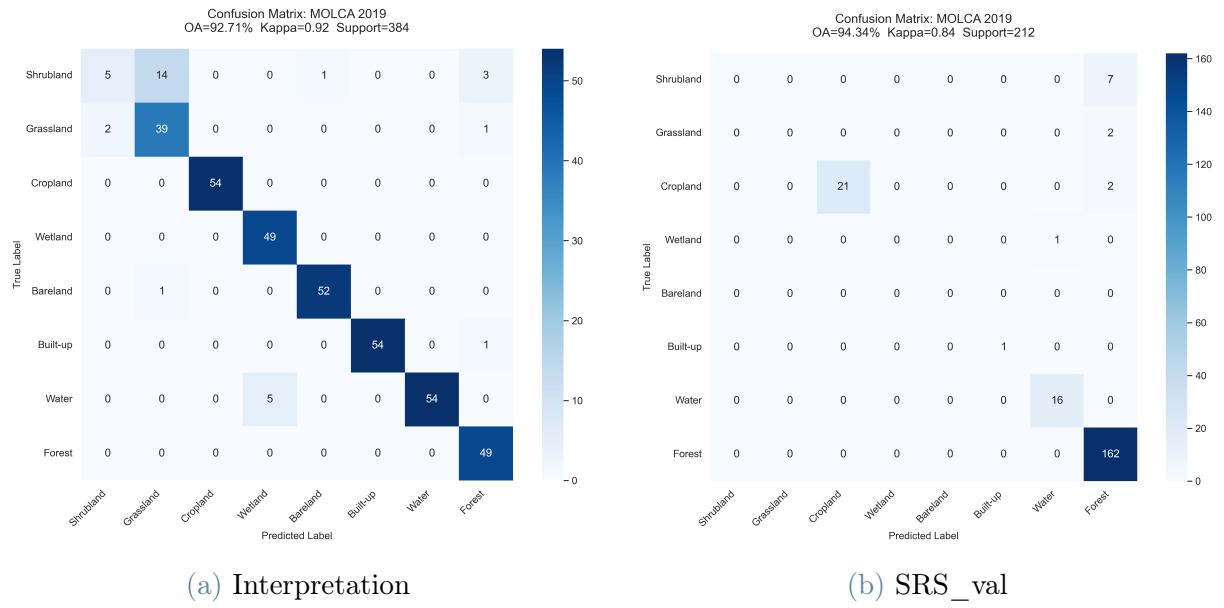


Figure 4.7: Confusion Matrices for MOLCA 2019

4.2. MOLCOPA

This section contains the analyses on the land cover maps derived from the MOLCOPA and the count map, focusing on the pros and cons of different derivation strategies.

4.2.1. Maps

The land cover maps derived from MOLCOPA by different strategies are shown in Figures 4.8, 4.9, 4.10, 4.11, 4.12, 4.13, 4.14 and 4.15.

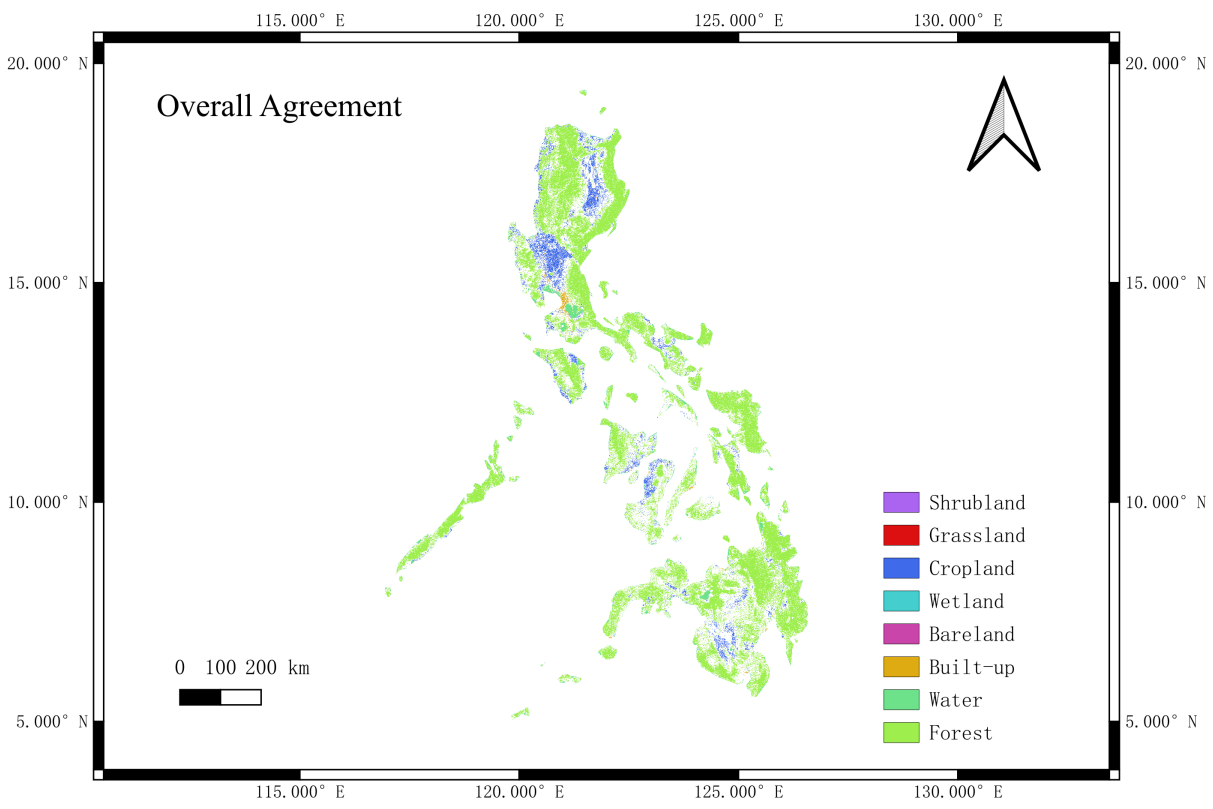


Figure 4.8: Overall Agreement

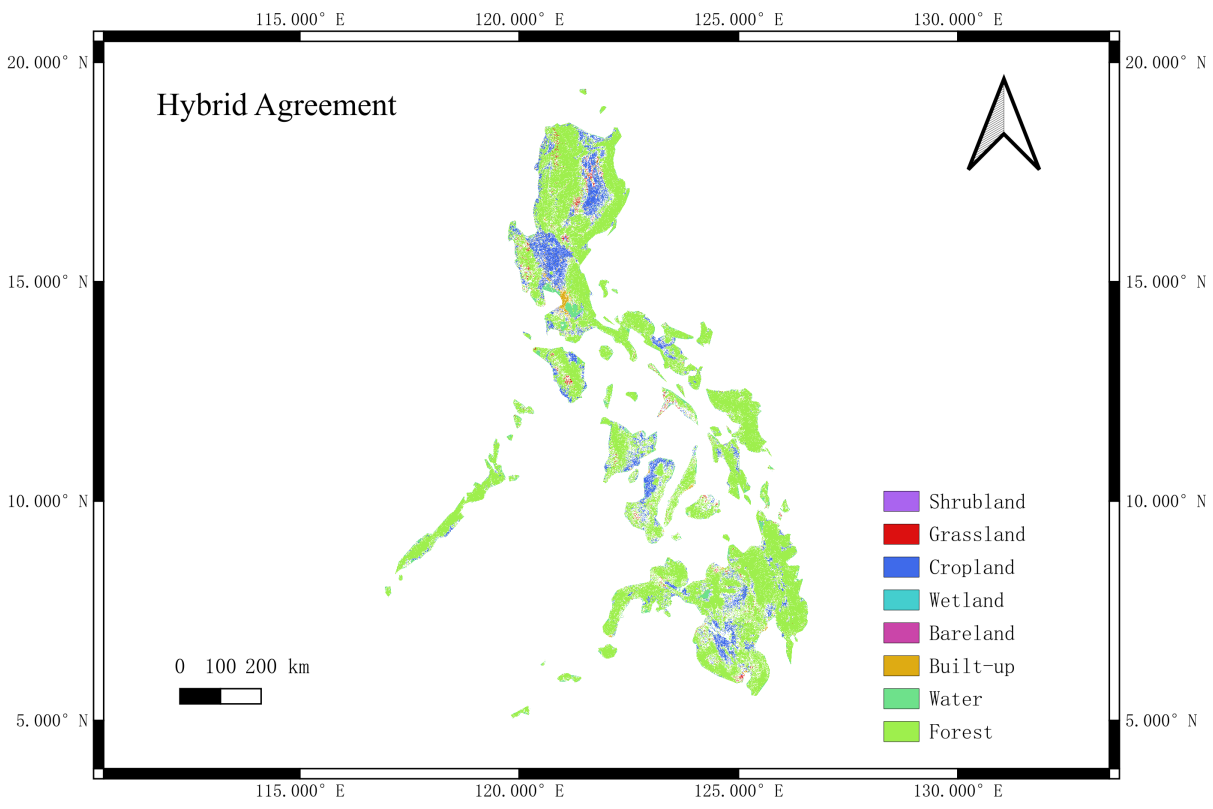


Figure 4.9: Hybrid Agreement

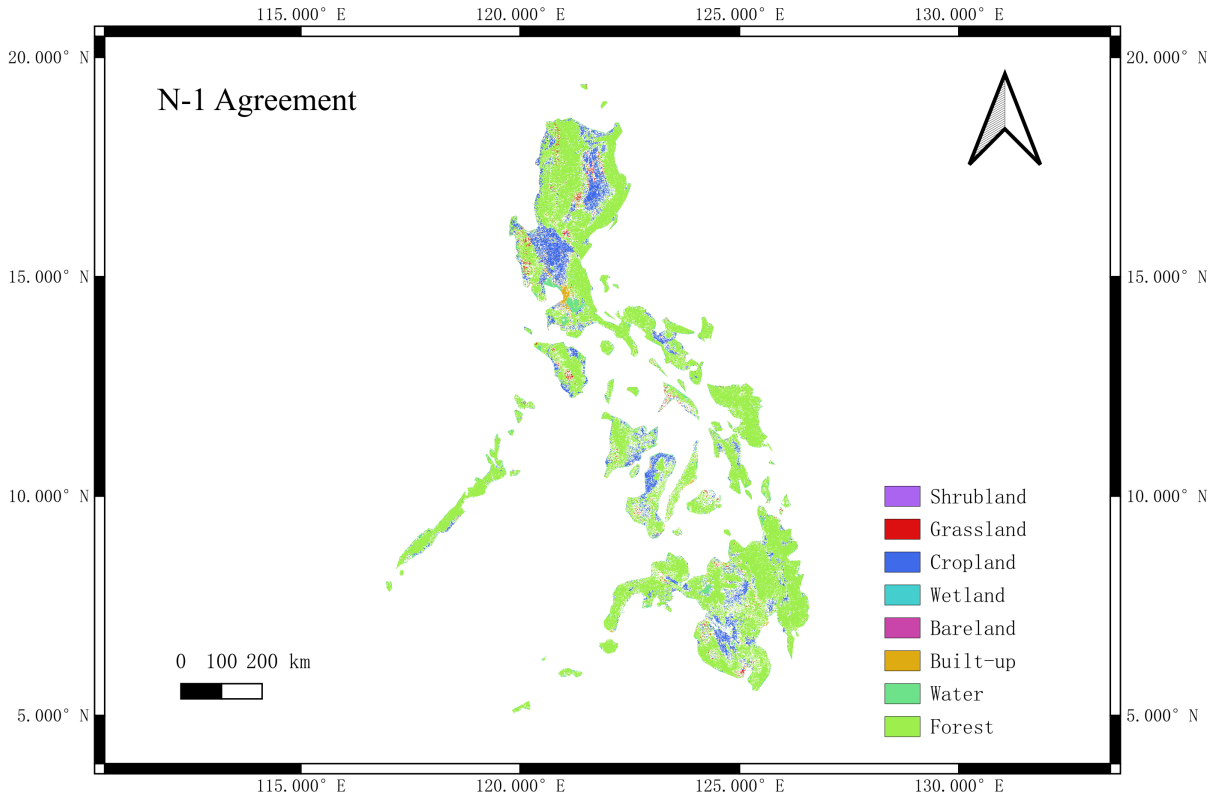


Figure 4.10: N-1 Agreement

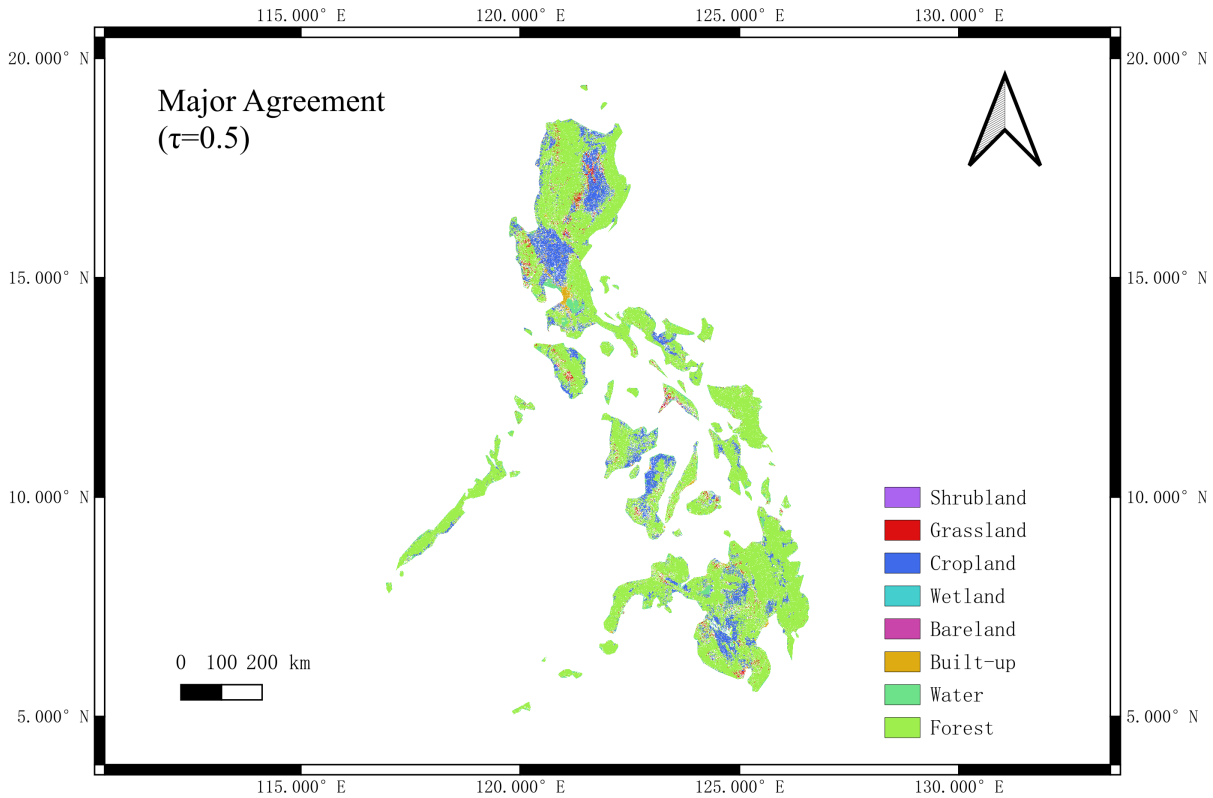


Figure 4.11: Major Agreement ($\tau = 0.5$)

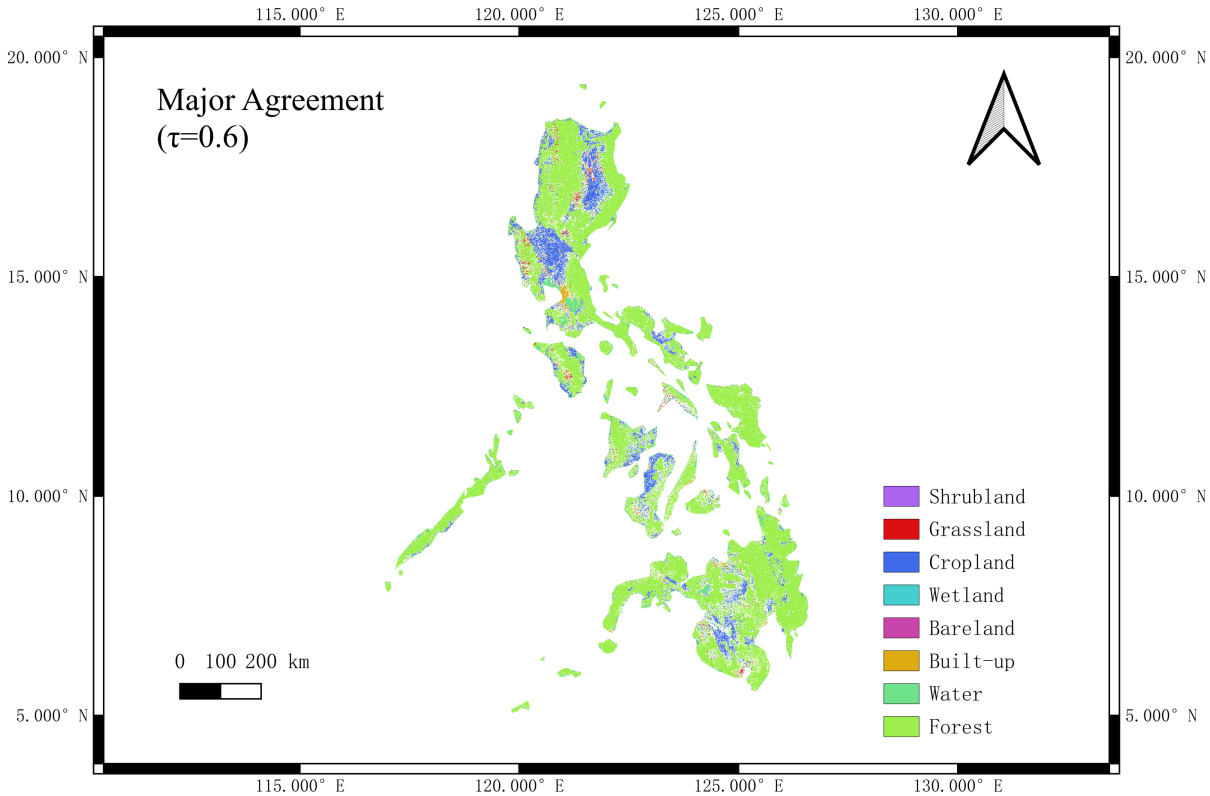


Figure 4.12: Major Agreement ($\tau = 0.6$)

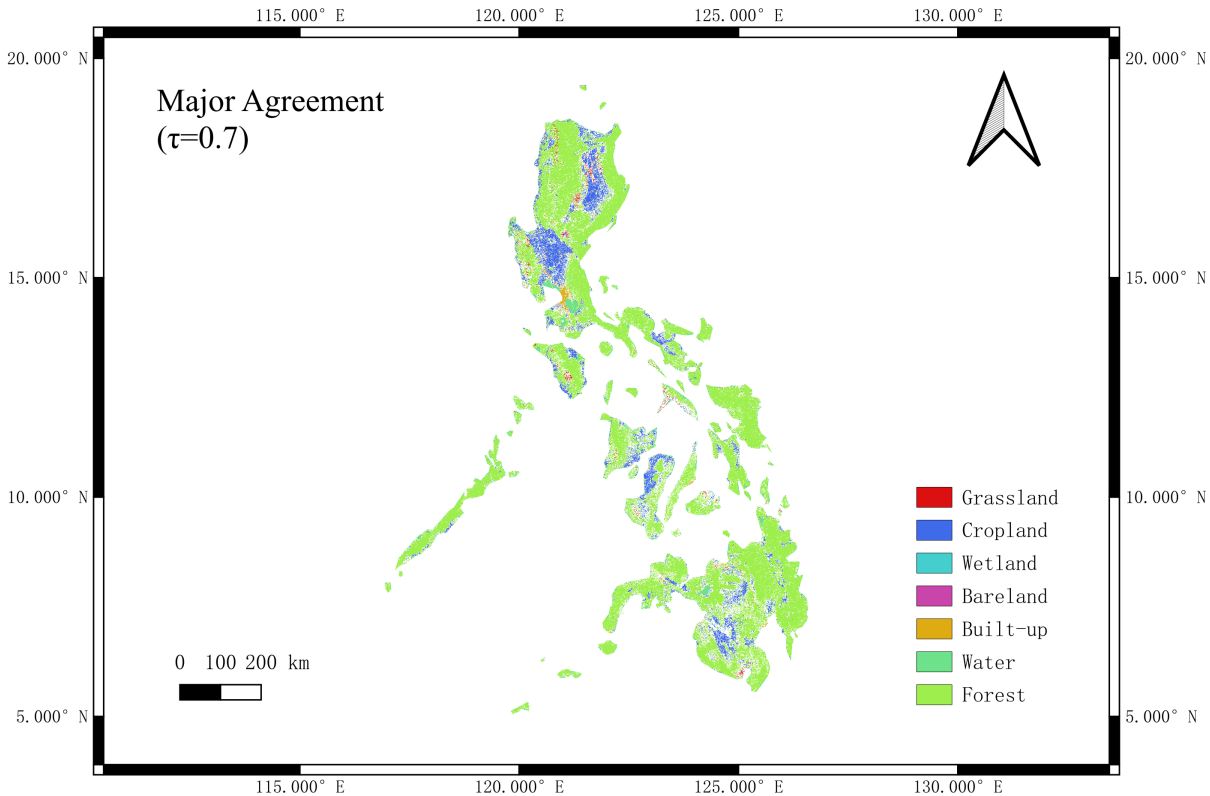


Figure 4.13: Major Agreement ($\tau = 0.7$)

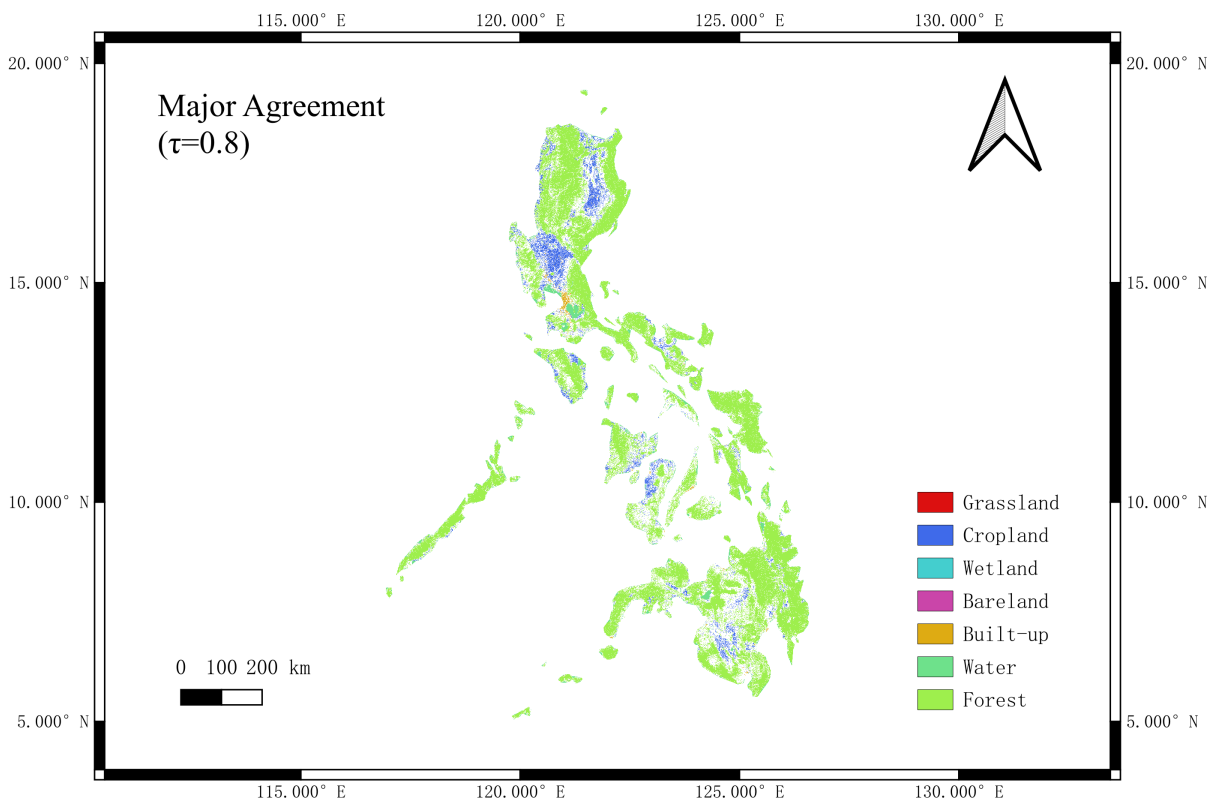


Figure 4.14: Major Agreement ($\tau = 0.8$)

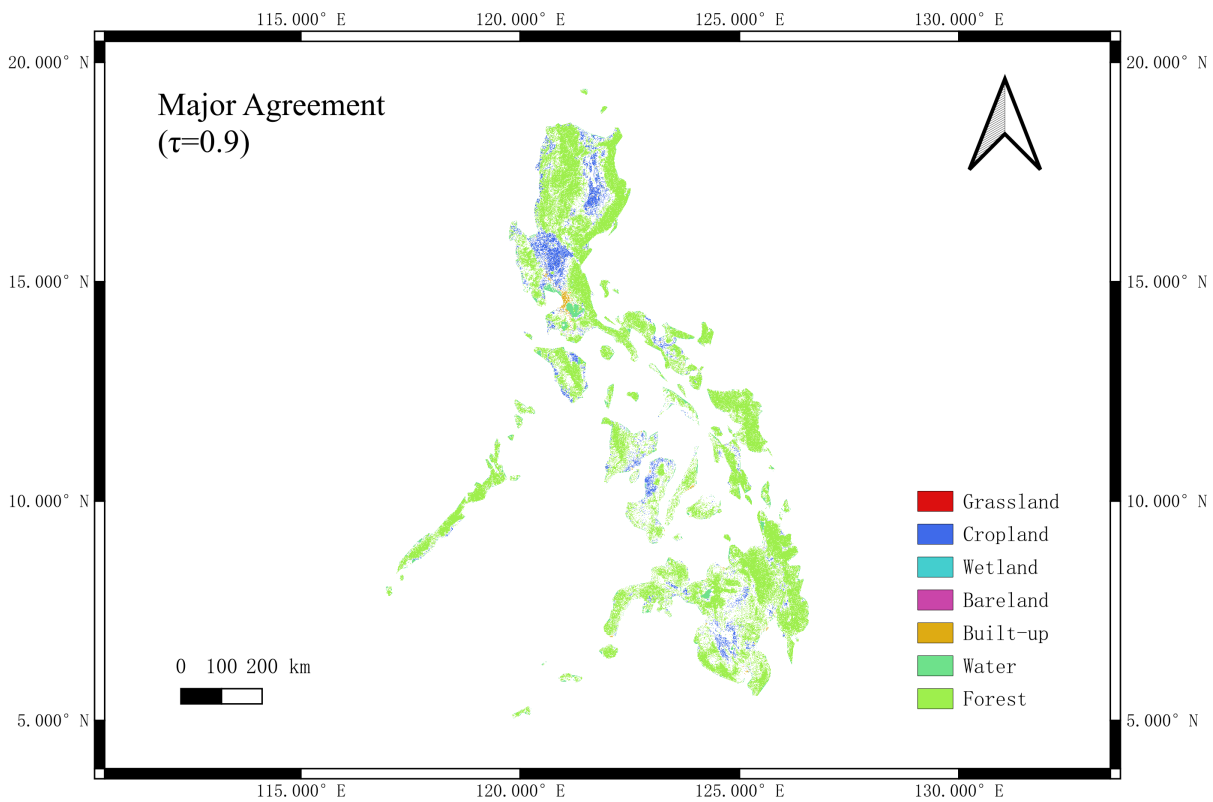


Figure 4.15: Major Agreement ($\tau = 0.9$)

4.2.2. Summary

Figure 4.16 shows the relationship between the coverage and overall accuracy of land cover maps under different land cover map derivation strategies.

The relationship between spatial coverage and Overall Accuracy reveals a distinct trade-off between data quantity and quality. As the derivation strategy becomes stricter, the Overall Accuracy significantly increases. However, this reliability comes at the cost of spatial coverage, leaving more of the study area as "NoData." In contrast, lenient strategies maximize coverage but introduce more noise. From the figure, it is believed that $\tau = 0.67$ can result in the most balanced results, which motivates the choice of the threshold for Hybrid Agreement.

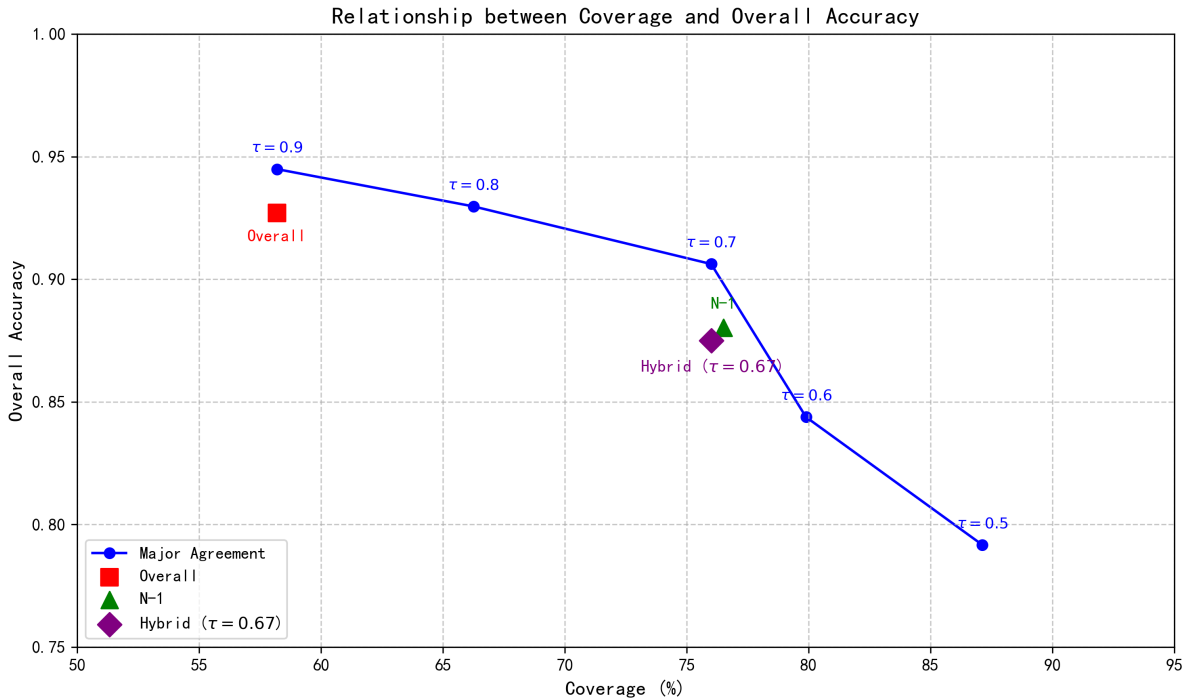


Figure 4.16: Relationship between Coverage and Overall Accuracy

4.2.3. Statistical Analysis

Tables 4.3–4.6 illustrate the pixel count and percentage statistics for land cover maps derived from MOLCOPA under different strategies. "Lichens and Mosses" and "Permanent ice and snow" are missing due to the fact that they do not exist in the study area.

The most prominent trend is the inverse relationship between the strictness of strategy and spatial coverage. It can be easily observed from the strategy of Major Agreement.

As the decision rule tightens from Major Agreement ($\tau = 0.5$) to Major Agreement ($\tau = 0.9$), the proportion of "NoData" pixels increases drastically from 12.88% to 41.81%. This indicates that a significant portion of the study area is characterized by conflicting information among the source products.

The impact of strict strategy is severe on minority classes, particularly "Shrubland" and "Grassland". "Shrubland" disappears under strict strategies. "Grassland" covers 1.93% of the area under the Major Agreement ($\tau = 0.5$). However, under the strict Major Agreement ($\tau = 0.9$), its coverage collapses to a negligible 0.0008% (only 22,902 pixels remaining out of 2.9 billion). This confirms that minority classes are rarely identified unanimously by all source products. The N-1 Agreement strategy proves to be a vital compromise here, maintaining a coverage of 1.02% for "Grassland" while keeping the "NoData" ratio (23.50%) much lower than the strict strategies.

Class	Overall	N-1	Hybrid ($\tau = 0.67$)	Major ($\tau = 0.5$)
Forest	1,423,007,516	1,770,645,803	1,767,201,809	1,943,233,068
No Data	1,215,388,342	683,075,010	697,846,379	374,555,566
Cropland	162,146,777	286,896,830	280,505,353	375,047,981
Water	91,150,155	107,597,884	107,592,398	116,150,917
Built-up	15,470,740	28,564,390	28,523,040	37,370,431
Grassland	26,738	29,587,628	24,731,060	56,061,405
Bareland	4,283	309,915	284,544	1,476,975
Wetland	3,732	517,904	513,700	3,299,021
Shrubland	7	2,926	7	2,926
Total	2,907,198,290	2,907,198,290	2,907,198,290	2,907,198,290

Table 4.3: Pixel Count Statistics for MOLCOPA-derived LC Maps (Part 1)

Class	Major ($\tau = 0.6$)	Major ($\tau = 0.7$)	Major ($\tau = 0.8$)	Major ($\tau = 0.9$)
Forest	1,847,184,656	1,767,199,513	1,643,640,176	1,423,005,204
No Data	584,538,625	697,863,119	981,161,485	1,215,405,157
Cropland	300,252,048	280,495,002	169,951,117	162,136,376
Water	113,032,747	107,592,364	96,500,352	91,150,121
Built-up	28,612,820	28,523,003	15,506,463	15,470,702
Grassland	30,413,731	24,727,232	22,902	22,902
Bareland	905,704	284,363	130,261	4,102
Wetland	2,255,033	513,694	285,534	3,726
Shrubland	2,926	0	0	0
Total	2,907,198,290	2,907,198,290	2,907,198,290	2,907,198,290

Table 4.4: Pixel Count Statistics for MOLCOPA-derived LC Maps (Part 2)

Class	Overall	N-1	Hybrid ($\tau = 0.67$)	Major ($\tau = 0.5$)
Forest	48.95%	60.91%	60.79%	66.84%
No Data	41.81%	23.50%	24.00%	12.88%
Cropland	5.58%	9.87%	9.65%	12.90%
Water	3.14%	3.70%	3.70%	4.00%
Built-up	0.53%	0.98%	0.98%	1.29%
Grassland	<0.01%	1.02%	0.85%	1.93%
Bareland	<0.01%	0.01%	0.01%	0.05%
Wetland	<0.01%	0.02%	0.02%	0.11%
Shrubland	<0.01%	<0.01%	<0.01%	<0.01%

Table 4.5: Percentage Statistics for MOLCOPA-derived LC Maps (Part 1)

Class	Major ($\tau = 0.6$)	Major ($\tau = 0.7$)	Major ($\tau = 0.8$)	Major ($\tau = 0.9$)
Forest	63.54%	60.79%	56.54%	48.95%
No Data	20.11%	24.00%	33.75%	41.81%
Cropland	10.33%	9.65%	5.85%	5.58%
Water	3.89%	3.70%	3.32%	3.14%
Built-up	0.98%	0.98%	0.53%	0.53%
Grassland	1.05%	0.85%	<0.01%	<0.01%
Bareland	0.03%	0.01%	<0.01%	<0.01%
Wetland	0.08%	0.02%	0.01%	<0.01%
Shrubland	<0.01%	0.00%	0.00%	0.00%

Table 4.6: Percentage Statistics for MOLCOPA-derived LC Maps (Part 2)

4.2.4. Accuracy Assessment

Accuracy assessment involves comparing the performance of different decision strategies using overall metrics (Accuracy and Kappa) and class-wise performance based on visual interpretation, and a consistency check against the external validation dataset (SRS_val).

Figure 4.17 illustrates the overall metrics (Overall Accuracy and Kappa Coefficient) for various decision strategies.

The "Major Agreement with $\tau = 0.9$ " strategy achieved the highest performance metrics (OA = 93.49%, Kappa = 0.924), slightly outperforming the "Overall Agreement" strategy (OA = 92.71%, Kappa = 0.916). However, the "N-1 Agreement" strategy, while achieving a slightly lower OA of 88.02% and Kappa of 0.863, demonstrated a robust balance, suggesting it handles conflicting data sources better than stricter thresholding methods.

A clear trend can be observed within the "Major Agreement" strategies. As the threshold (τ) increases from 0.5 to 0.9, the Overall Accuracy improves significantly from 79.17% to 93.49%, and the Kappa coefficient rises from 0.762 to 0.924. This confirms that stricter agreement constraints effectively filter out ambiguous pixels, resulting in higher confidence predictions.

Figure 4.18 shows the Overall Accuracy results based on visual interpretation and SRS_val. Although there are gaps between two approaches, the trend is highly consistent, which improves the reliability of the results by visual interpretation.

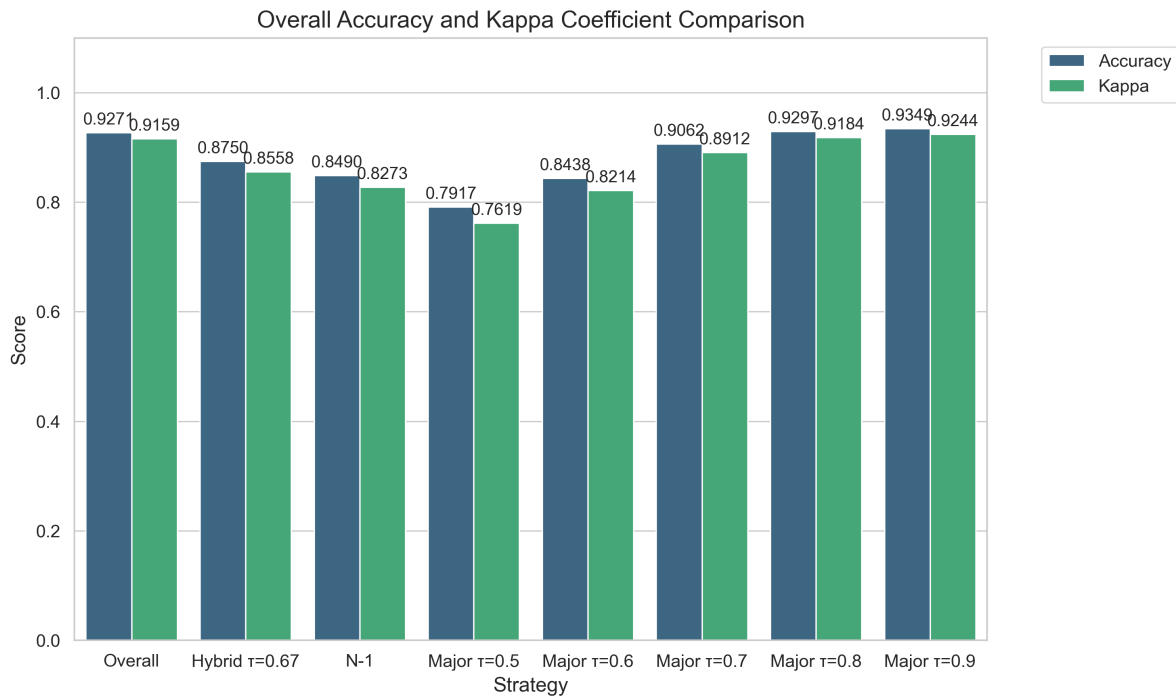


Figure 4.17: Overall Accuracy and Kappa Coefficient Comparison (Interpretation)

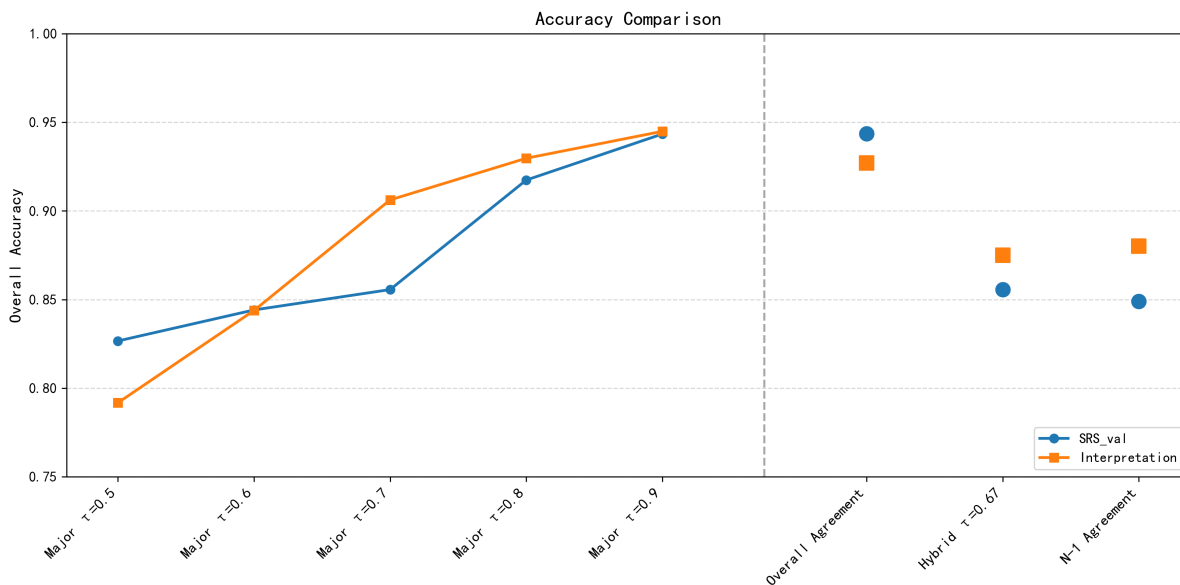


Figure 4.18: Accuracy Comparison by Visual Interpretation and SRS_val

Figures 4.19–4.21 illustrate the class-wise metrics (recall, precision, and F1-score) based on visual interpretation. Figures 4.22–4.29 show the confusion matrices derived from visual interpretation and validation against SRS_val.

The class-wise metrics reveal distinct performance variations across land cover classes. Classes such as "Cropland," "Water," and "Built-up," exhibit consistently high performance across all strategies. For instance, in the "Overall Agreement" strategy, Cropland achieved a perfect F1-score of 1.00, and Built-up reached 0.99, indicating that MOLCOPA is highly effective in identifying spectrally distinct features.

The "Shrubland" proves to be the most difficult to classify, revealing a critical trade-off. The scores of major agreement with $\tau \geq 0.7$ are 0.00 as no pixel is "Shrubland" on the maps. For other strategies, the performances are typically much lower than other classes.

It can be observed from the confusion matrices that vegetation classes are prone to confusion (particularly shrubland, grassland, and forest). "Shrubland" is frequently misclassified as grassland or forest. In contrast, cropland is relatively the most distinguishable among the vegetation categories. Additionally, water is prone to be misclassified as wetland. Overall, bareland, built-up, and water are the best-performing classes.

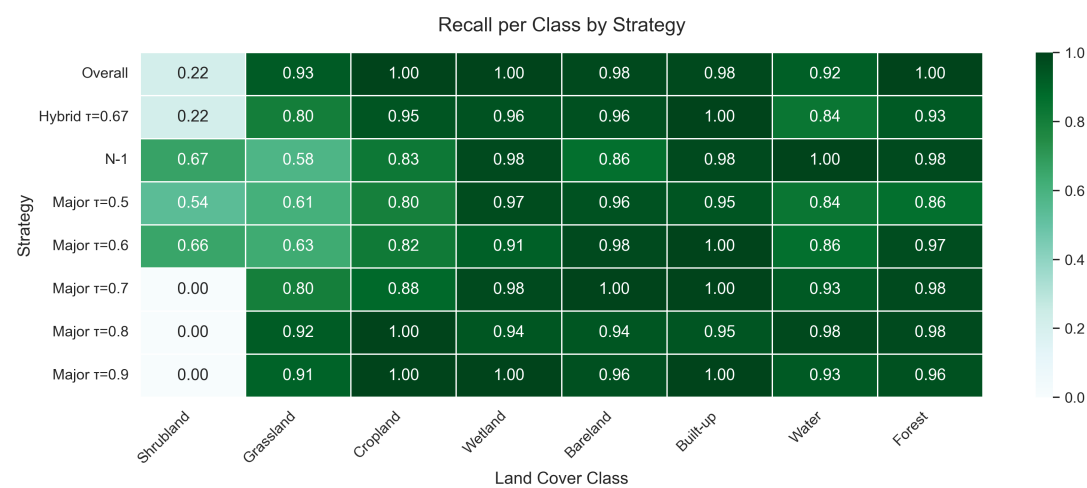


Figure 4.19: Recall per Class by Strategy (Interpretation)

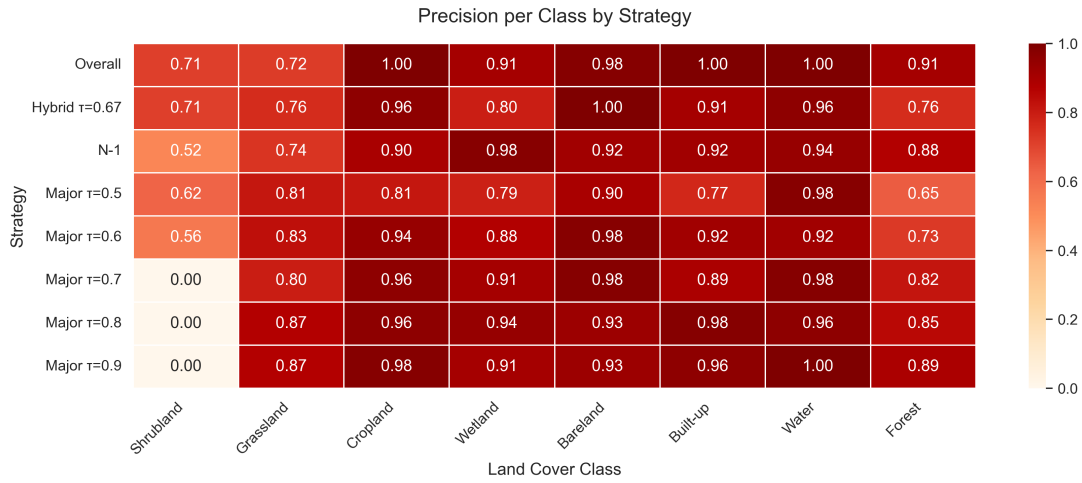


Figure 4.20: Precision per Class by Strategy (Interpretation)

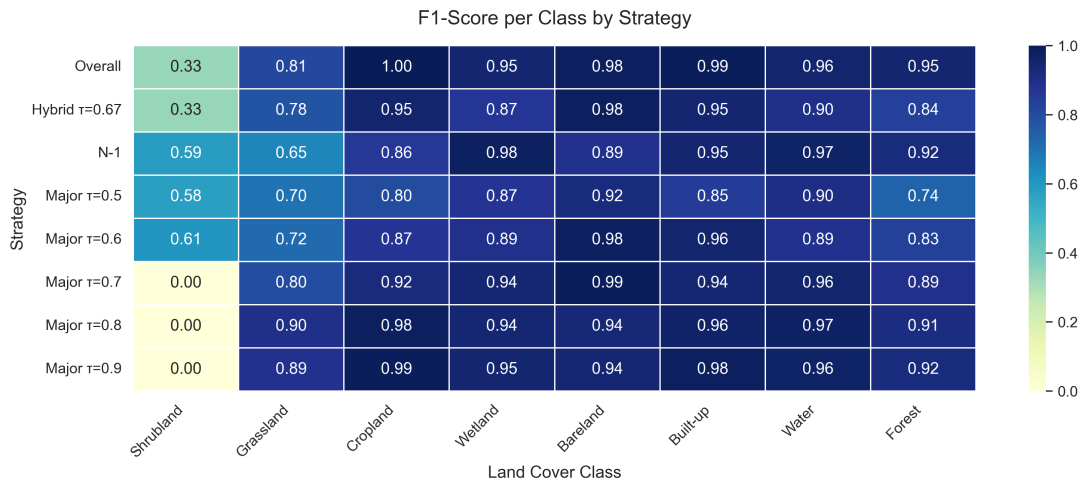


Figure 4.21: F1-score per Class by Strategy (Interpretation)

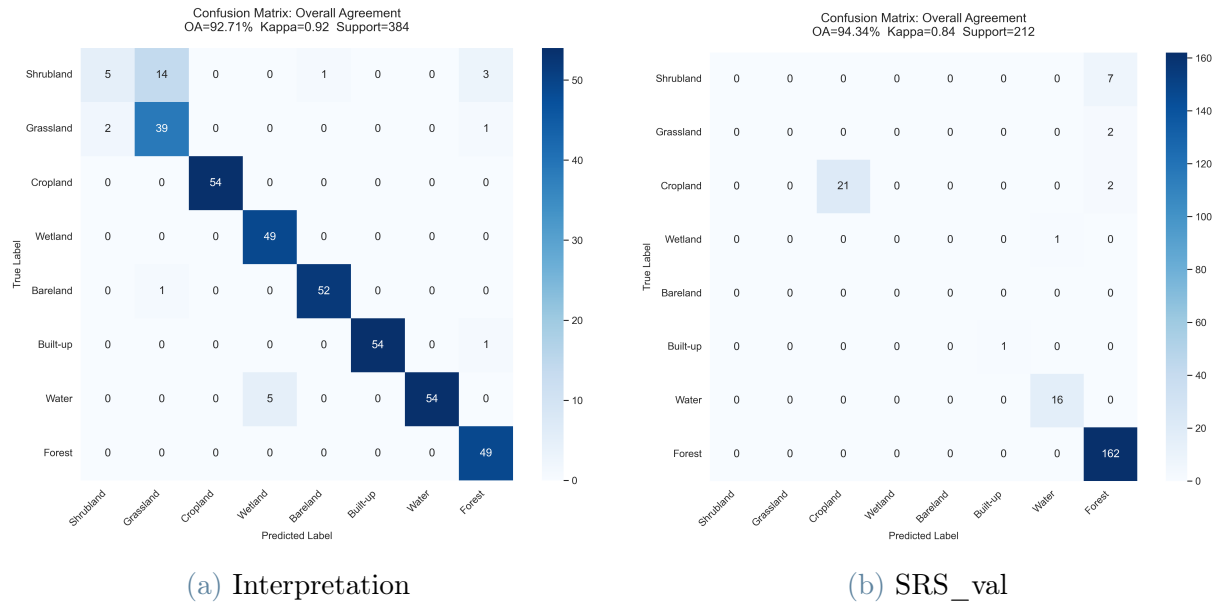


Figure 4.22: Confusion Matrices for Overall Agreement

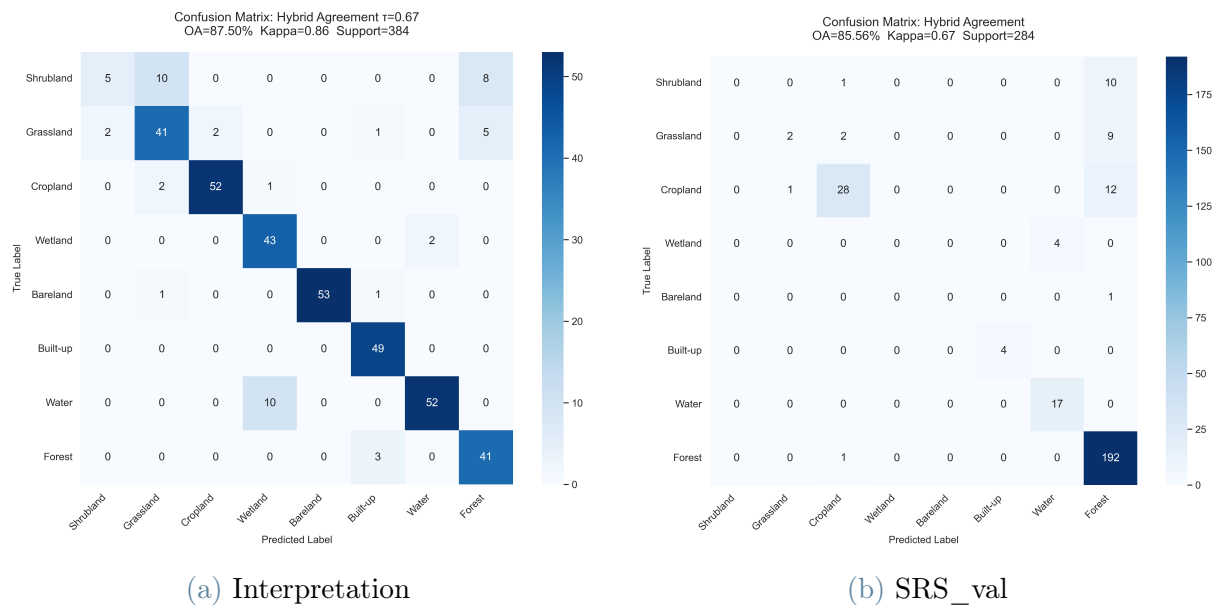


Figure 4.23: Confusion Matrices for Hybrid Agreement

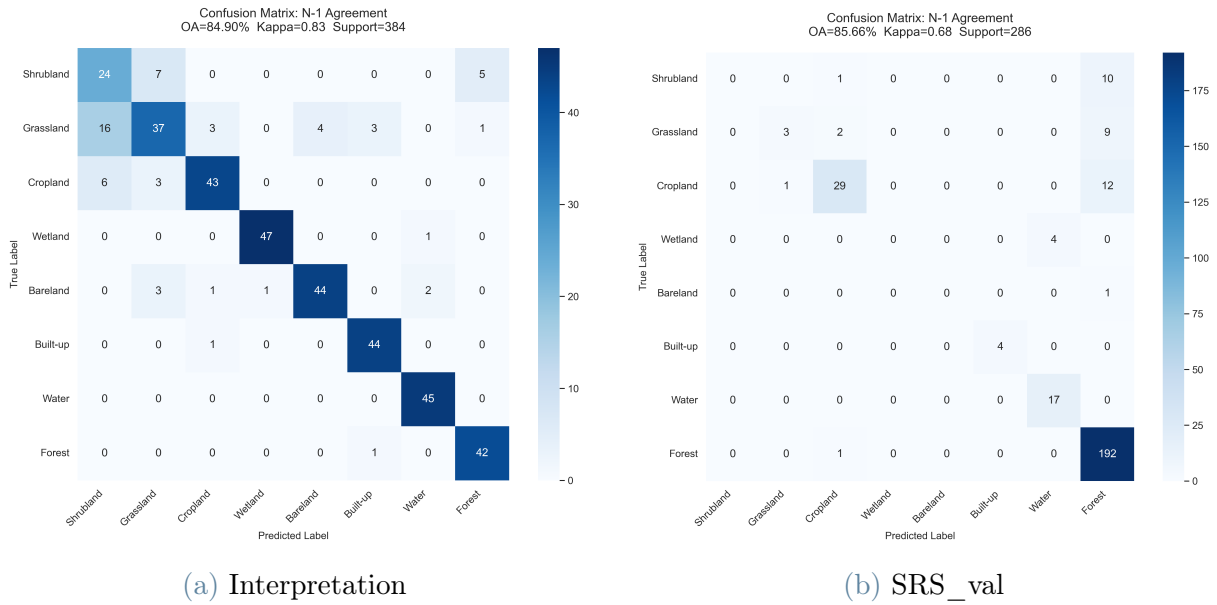
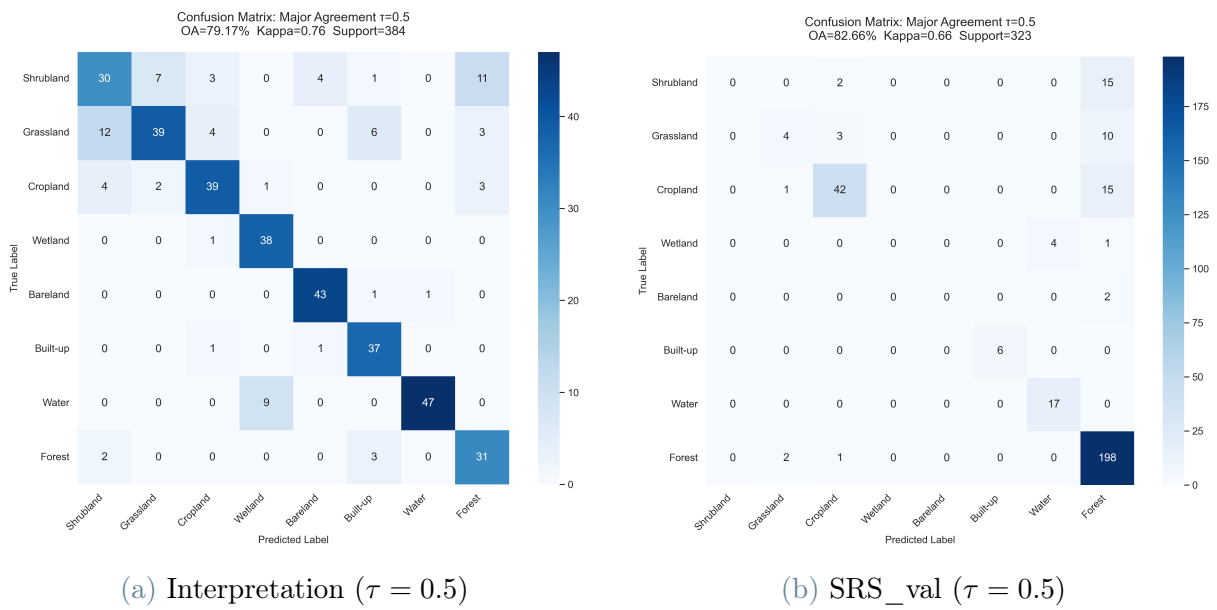
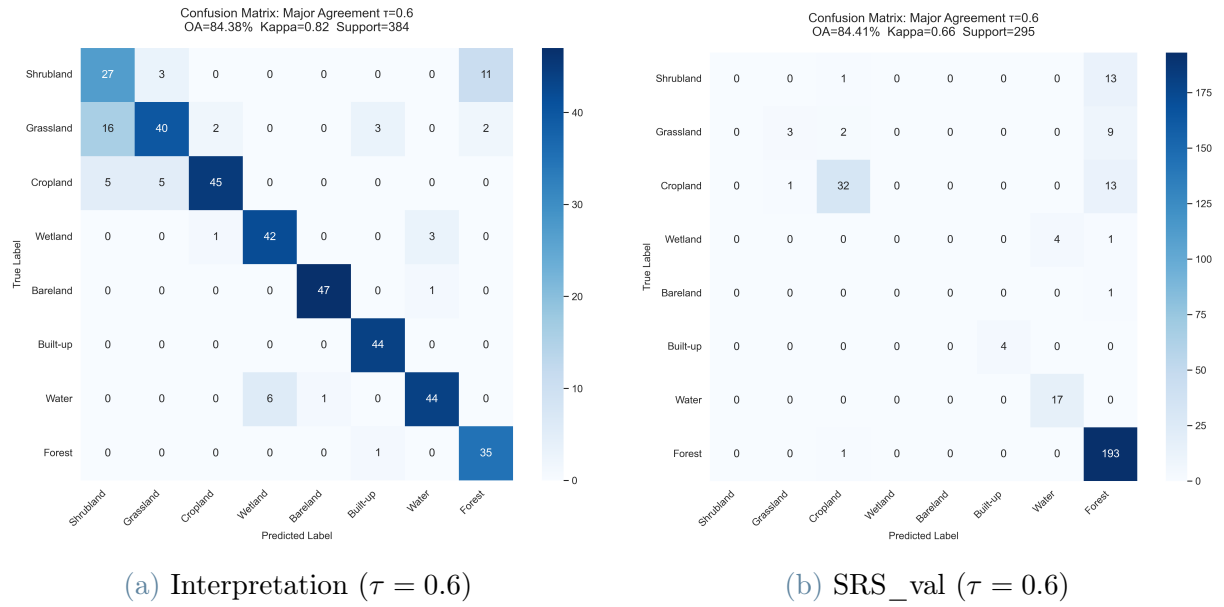
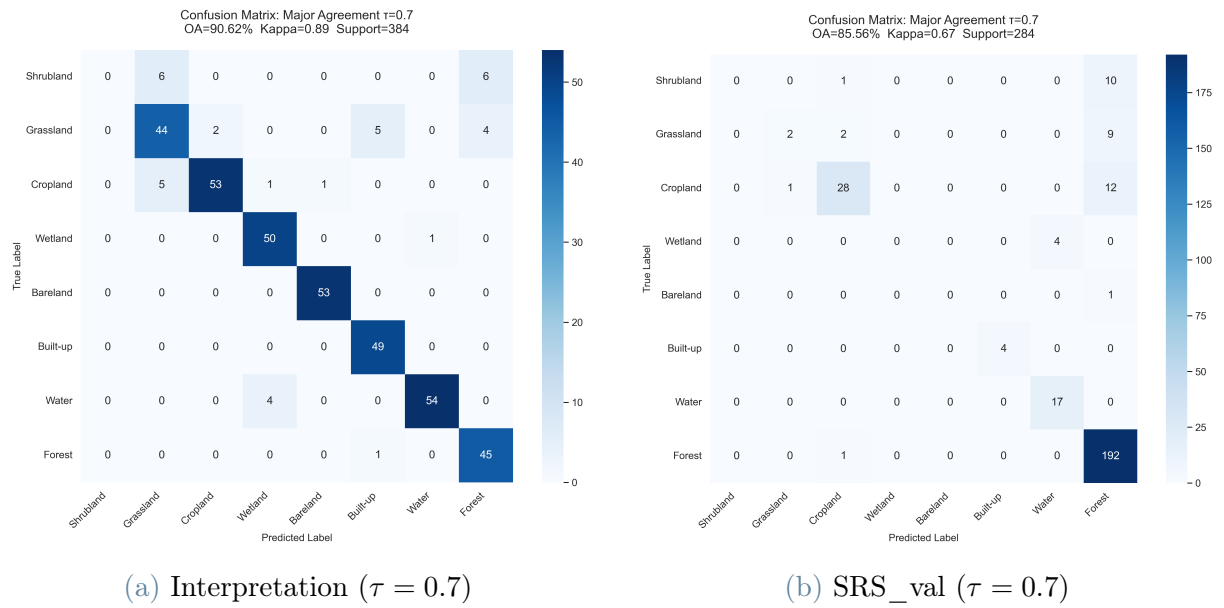
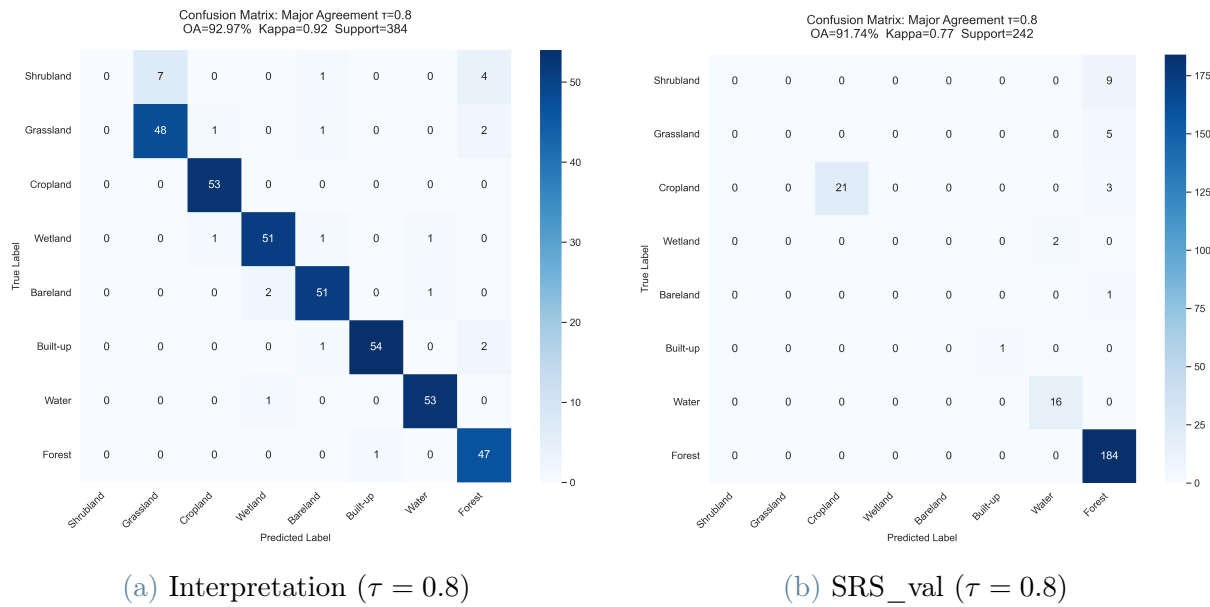
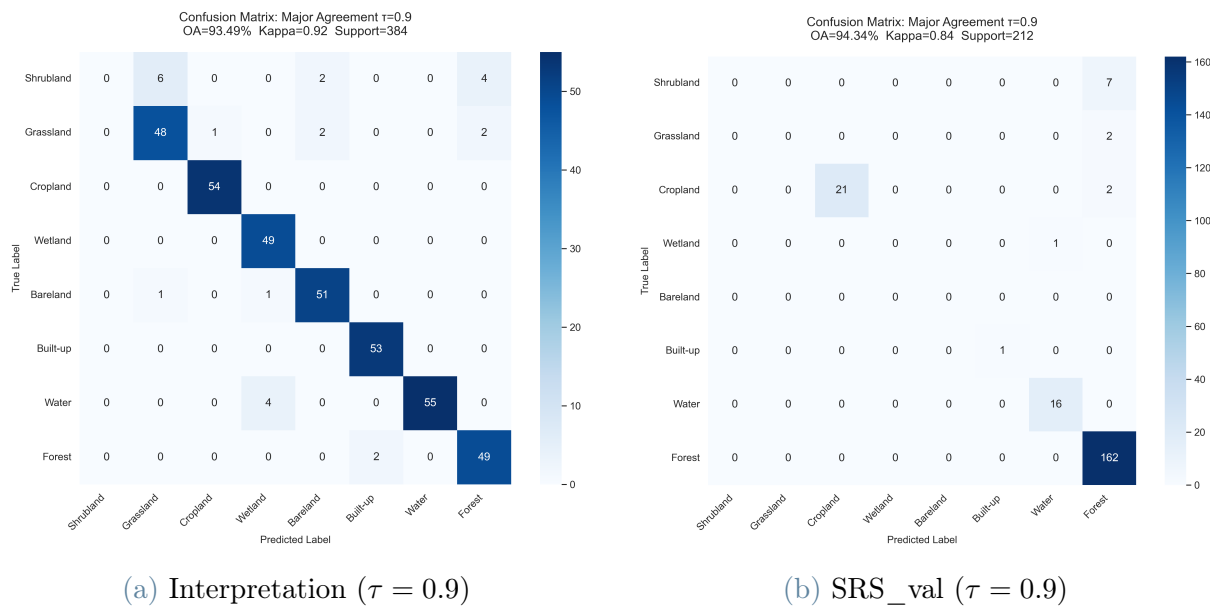


Figure 4.24: Confusion Matrices for N-1 Agreement

Figure 4.25: Confusion Matrices for Major Agreement at $\tau = 0.5$

Figure 4.26: Confusion Matrices for Major Agreement at $\tau = 0.6$ Figure 4.27: Confusion Matrices for Major Agreement at $\tau = 0.7$

Figure 4.28: Confusion Matrices for Major Agreement at $\tau = 0.8$ Figure 4.29: Confusion Matrices for Major Agreement at $\tau = 0.9$

4.3. GeoAI-based Land Cover Mapping

This section includes the presentation of land cover maps generated based on GeoAI and their statistical and performance analysis.

4.3.1. Unsupervised Approach

In this part, the results of the unsupervised approach are presented, including maps, statistical and performance analysis.

4.3.1.1. Maps

The land cover maps generated by different strategies are shown in Figures 4.30, 4.31, and 4.32.

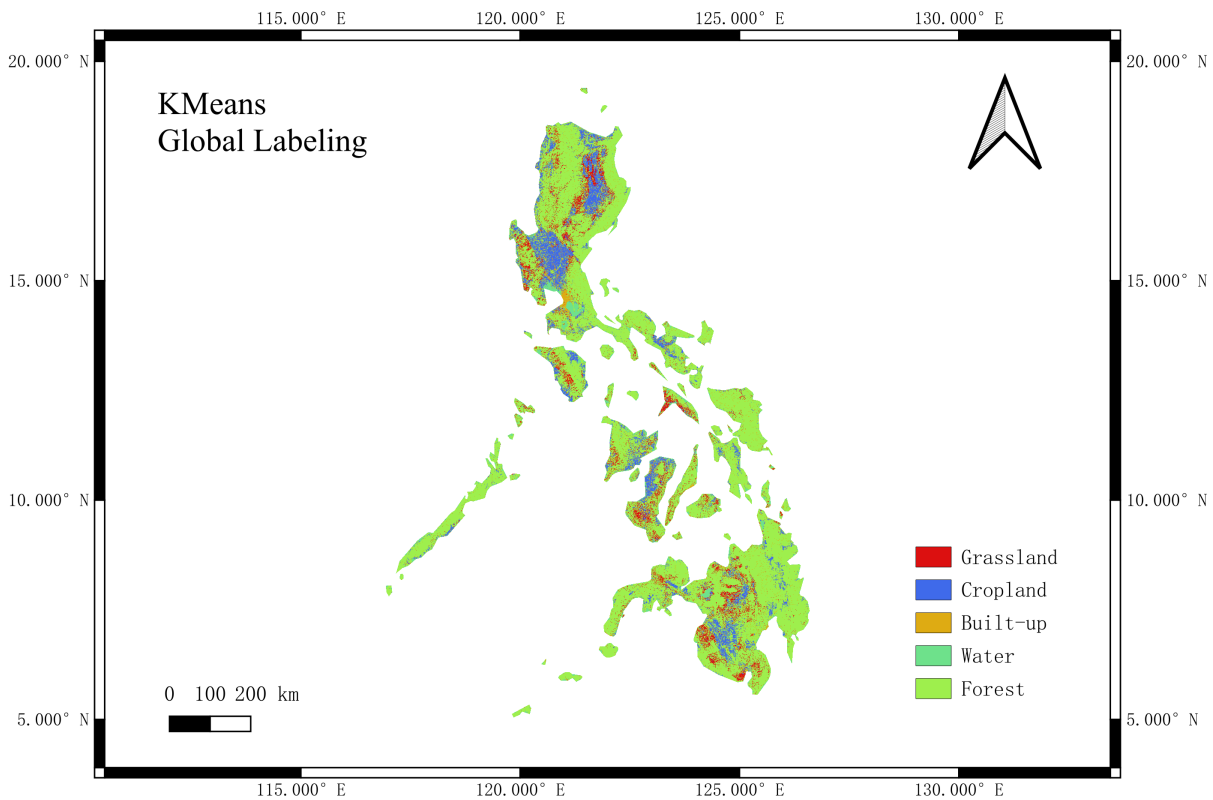


Figure 4.30: Land Cover Map by KMeans with Global Labeling

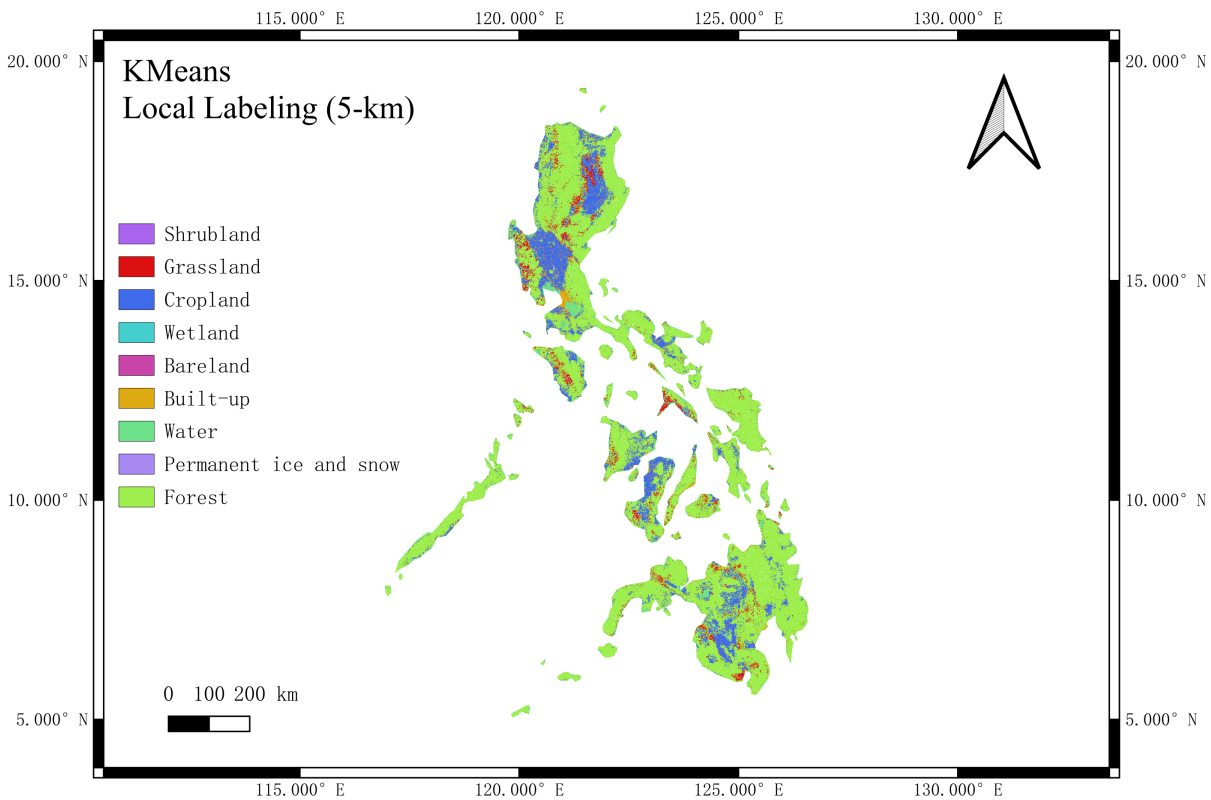


Figure 4.31: Land Cover Map by KMeans with Local Labeling (5-km)

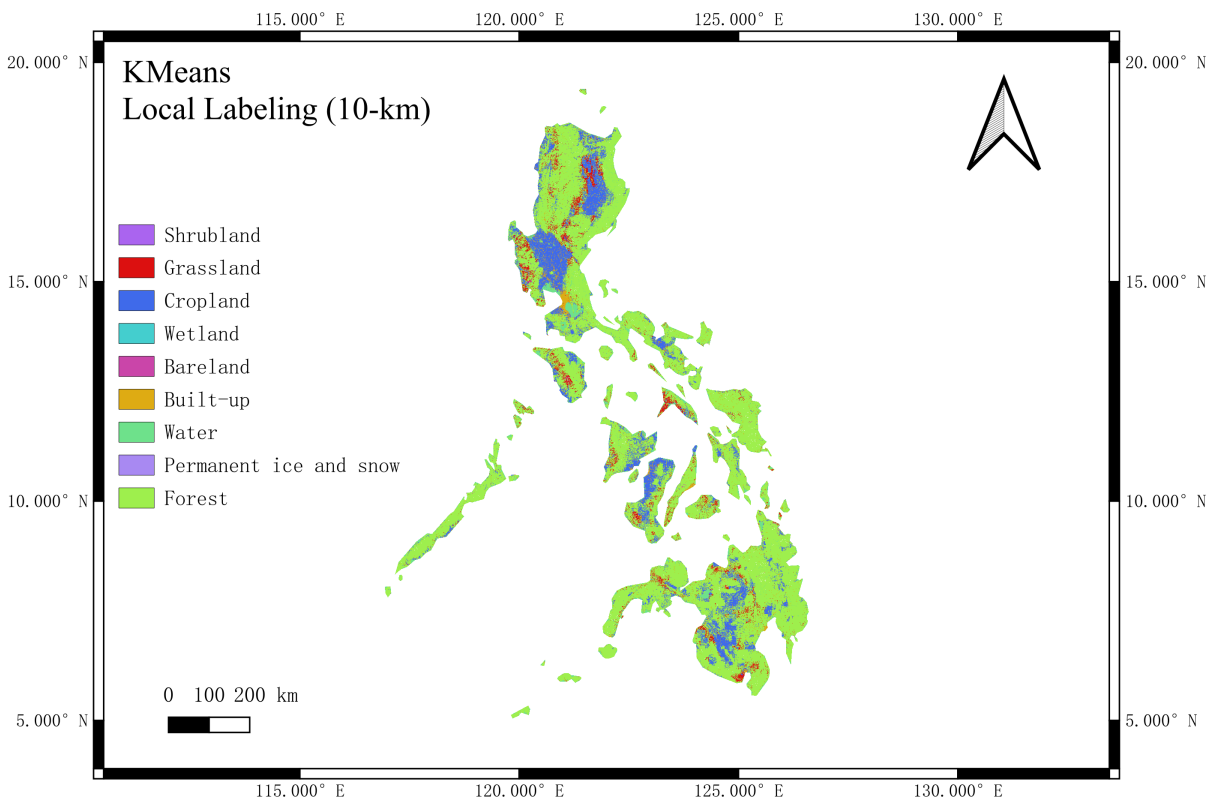


Figure 4.32: Land Cover Map by KMeans with Local Labeling (10-km)

4.3.1.2. Statistical Analysis

Tables 4.7 and 4.8 present a quantitative comparison of land cover maps generated using three different automatic labeling strategies: Global Labeling, Local Labeling (5-km) and Local Labeling (10-km).

The Global Labeling strategy exhibits a strong bias towards dominant classes, resulting in a simplified representation of the landscape. It assigns 100% of the pixels to only five major classes ("Forest," "Cropland," "Grassland," "Water," and "Built-up"). "Forest" alone occupies 72.58% of the area. Crucially, this strategy completely fails to identify spatially rare classes. "Bareland", "Wetland", "Shrubland", and "Permanent ice and snow" all record a pixel count of 0. This occurs because Global Labeling aggregates class probabilities across the entire study area.

The introduction of Local Labeling (both 5-km and 10-km) resolves the issue of missing classes by allowing for spatially adaptive class assignment. Both local strategies successfully recover the missing classes. Meanwhile, some pixels are labeled "NoData" because the MOLCOPA is not complete (caused by the binary constraints); if all the pixels belonging to a specific cluster have "NoData" in the local window, the pixels cannot be labeled correctly, resulting in the "NoData" pixels in the final land cover map. Dramatically, "Permanent ice and snow" pixels appear in the land cover maps under local labeling strategy, resulting from the "Permanent ice and snow" pixels from Esri LULC and Dynamic World as illustrated in Section 2.7.

Class	Global Labeling	Local Labeling (5-km)	Local Labeling (10-km)
Forest	2,172,113,139	2,143,272,001	2,146,407,091
Cropland	333,827,952	453,771,651	448,912,153
Grassland	235,781,921	137,582,269	141,899,473
Water	126,103,301	132,865,360	134,426,171
Built-up	124,955,390	97,463,961	110,313,165
Bareland	0	5,068,245	4,875,954
No Data	0	19,866,085	4,870,158
Wetland	0	2,497,333	899,463
Shrubland	0	386,290	164,859
Permanent ice and snow	0	8,508	13,216
Total	2,992,781,703	2,992,781,703	2,992,781,703

Table 4.7: Pixel Count Statistics for Land Covers Maps by Unsupervised Approach

Class	Global Labeling	Local Labeling (5-km)	Local Labeling (10-km)
Forest	72.5784%	71.6147%	71.7195%
Cropland	11.1544%	15.1622%	14.9998%
Grassland	7.8784%	4.5971%	4.7414%
Water	4.2136%	4.4395%	4.4917%
Built-up	4.1752%	3.2566%	3.6860%
Bareland	0.0000%	0.1693%	0.1629%
No Data	0.0000%	0.6638%	0.1627%
Wetland	0.0000%	0.0834%	0.0301%
Shrubland	0.0000%	0.0129%	0.0055%
Permanent ice and snow	0.0000%	0.0003%	0.0004%

Table 4.8: Percentage Statistics for Land Covers Maps by Unsupervised Approach

4.3.1.3. Accuracy Assessment

Further accuracy analyses are based on the maps produced under 2 local labeling strategies. The map under global labeling strategy is not included due to the terrible performance in identifying minor classes in the study area.

Figures 4.33 and 4.34 show the confusion matrices, Overall Accuracy and Kappa Coefficient on ValiSet-O (including its two subsets) and SRS_val for 2 local labeling strategies.

Local labeling with a 5-km $\times$ 5-km window size has better performance on all validation datasets. It is worth noting that they have comparable performance on SRS_val but show different performances on ValiSet-O which is considered more reliable and robust. The excellent performance on SRS_val is attributed to its imbalanced classes (major classes are predominant), which will lead to an overestimation in its performance.

Across ValiSet-O and its two subsets, the 5-km window size consistently outperforms 10-km. Furthermore, both maps perform significantly better in the "easy" area compared to the "hard" area, with a huge performance gap of approximately 20%. However, results on the full ValiSet-O (all samples) remain unsatisfactory, achieving Overall Accuracies of 56.12% and 61.07%, for 10-km and 5-km window size respectively.

Unsupervised approach generally struggles with minority classes, but the 5 km window size offers substantial improvement. With the 10-km window size, classes such as "Wetland," "Bareland," and "Shrubland" are nearly impossible to classify correctly; however, the 5 km window yields a drastic performance boost for these specific categories. The primary challenge for the unsupervised method lies in the classification of minority classes, which are frequently mislabeled as major classes.

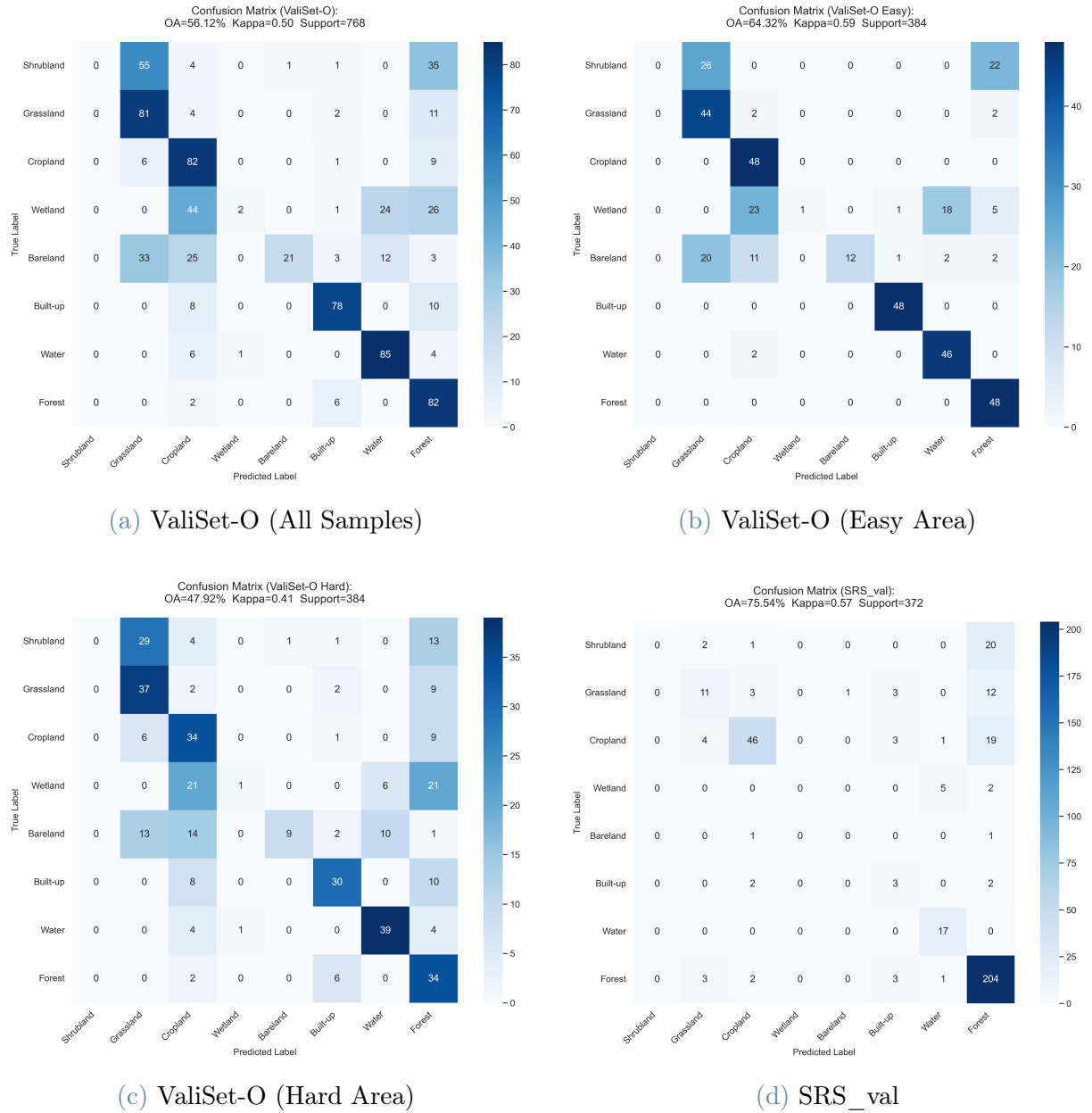


Figure 4.33: Confusion Matrices for Local Labeling (10-km)

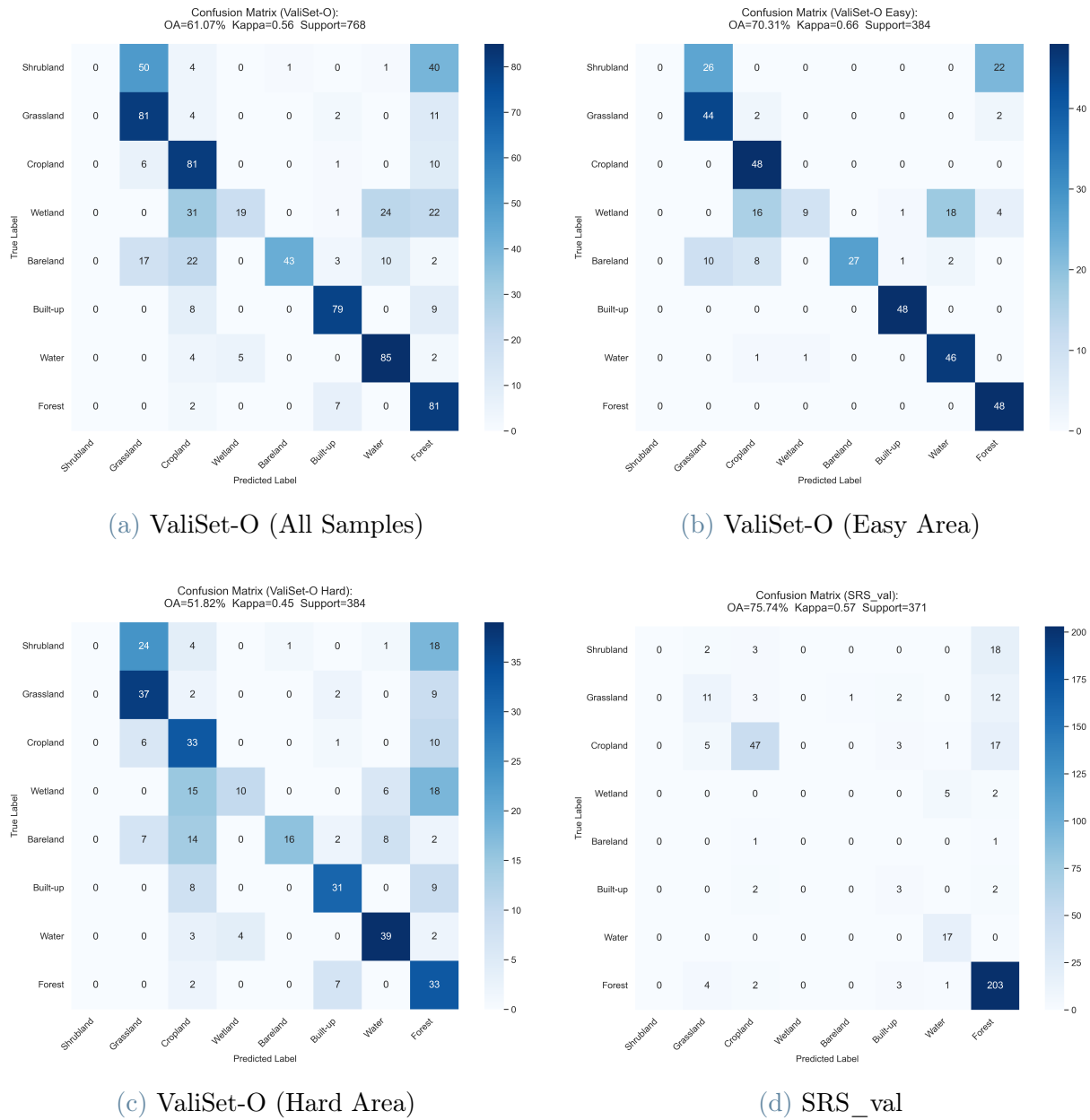


Figure 4.34: Confusion Matrices for Local Labeling (5-km)

4.3.2. Semi-supervised Approach

In this part, the accuracy assessment for all the experiment groups will be included, illustrating the optimal configurations for three classifiers and the impact of experimental variables. Land Cover Maps generated by three classifiers with their optimal configurations will be further analyzed with detailed confusion matrices and statistics.

4.3.2.1. Preliminary Experiments

In this part, the results of the designed preliminary experiments are presented. The overall performance analysis in this part is based on ValiSet-O, which is more robust and reliable than SRS_val. Table 4.9 illustrates the Overall Accuracy and Kappa Coefficient on ValiSet-O and SRS_val for all the designed experiment groups. The final MLP performance is determined by selecting the best result among five separate training runs.

1. Impact of Features: Full Stack vs. Embeddings Only

The most significant finding from the experiment is the consistent superiority of the Full Stack configuration over the Embeddings Only baseline. Across all classifiers and sample sizes, the integration of MOLCOPA probabilities and the Count map yields a performance improvement of approximately 3.5% to 5.5% against ValiSet-O in Overall Accuracy. For instance, the RF classifier (Set-50) improves from an accuracy of 74.48% (Embeddings Only) to 80.34% (Full Stack) against ValiSet-O. Similarly, the Kappa coefficient increases from 0.7083 to 0.7753.

This result also strongly refutes the possibility that the model learns only from the MOLCOPA probabilities and the count map, in which case it will have a terrible performance on ValiSet-O. Instead, it demonstrates that the "Full Stack" effectively combines the semantic information of embeddings with the statistical confidence of the MOLCOPA.

2. Classifier Comparison

Comparing the three algorithms, SVM and MLP demonstrated slightly higher robustness compared to RF, particularly with "Embeddings Only". Interestingly, when the MOLCOPA and count map are added, the performance gap narrows, with all three classifiers converging to the 79-81% range. This indicates that RF benefits significantly from the MOLCOPA and the count map, although MLP and SVM outperform RF.

3. Impact of Training Sample Size

The experiment reveals high data efficiency in the proposed framework. Increasing the sample size from 50 to 200 pixels per class does not result in a linear performance increase.

Notably, it is counter-intuitive that RF performance actually declines slightly as sample size increases (from 80.34% at Set-50 to 79.82% at Set-200). No classifier shows a constant increase in accuracy with sample size, demonstrating its stability. For MLP, the accuracy peaked at Set-100 (81.12%) and slightly decreased at Set-200 (80.86%). However, the counter-intuitive performance of RF and SVM may be caused by SMOTE as described in Section 3.5.2, which over-sampled "Shrubland" to the same number of any other class. Providing a balanced training dataset, the results may be different.

Classifier	Dataset	Embeddings Only				Full Stack			
		ValiSet-O		SRS		ValiSet-O		SRS	
		OA	Kappa	OA	Kappa	OA	Kappa	OA	Kappa
RF	Set-50	74.48	0.7083	76.88	0.6031	80.34	0.7753	77.69	0.8034
	Set-100	74.61	0.7098	77.15	0.6073	79.95	0.7708	77.15	0.6068
	Set-200	74.48	0.7038	77.42	0.6108	79.82	0.7694	77.69	0.6139
SVM	Set-50	76.95	0.7366	75.00	0.5797	80.47	0.7768	77.15	0.6050
	Set-100	76.69	0.7336	76.34	0.5979	79.82	0.7694	75.54	0.5805
	Set-200	76.82	0.7351	76.08	0.5873	80.34	0.7753	77.42	0.6082
MLP	Set-50	75.78	0.7232	75.00	0.5916	80.21	0.7738	77.15	0.6145
	Set-100	76.82	0.7351	76.34	0.6023	81.12	0.7843	75.81	0.5861
	Set-200	77.21	0.7396	75.00	0.5849	80.86	0.7813	77.42	0.6134

Table 4.9: Comparison of Overall Accuracy and Kappa Coefficient on ValiSet-O and SRS datasets

4.3.2.2. Maps

Figures 4.35, 4.36, and 4.38 show the land cover maps generated by RF, SVM, and MLP with their optimal configurations from the perspective of Overall Accuracy against ValiSet-O (RF-50B, SVM-50B and MLP-100B, respectively), in which 3 land cover maps show similar results. Furthermore, maps for the MLP-50B and MLP-200B configurations were also generated (Figure 4.37 and 4.39) due to the discrepancy in the overall performance between ValiSet-O and SRS_val alongside a reason discussed in a later section.

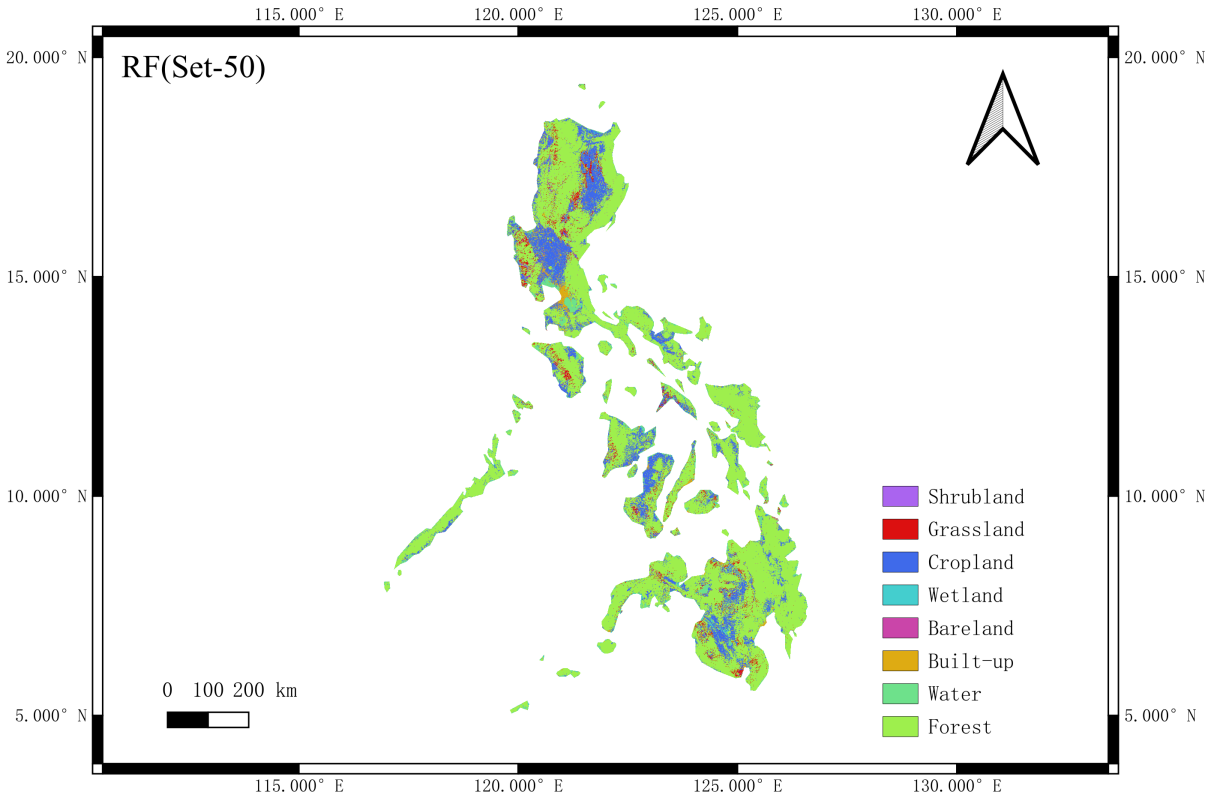


Figure 4.35: Land Cover Map by RF with Set-50

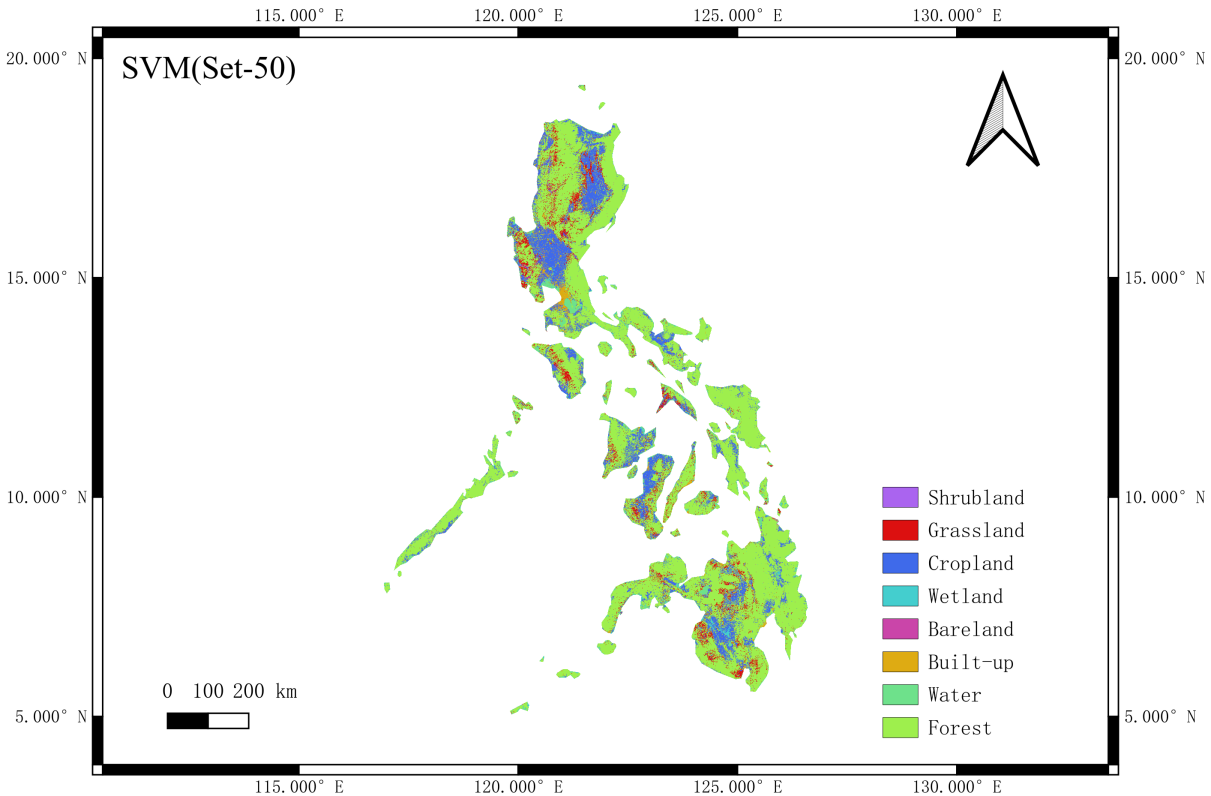


Figure 4.36: Land Cover Map by SVM with Set-50

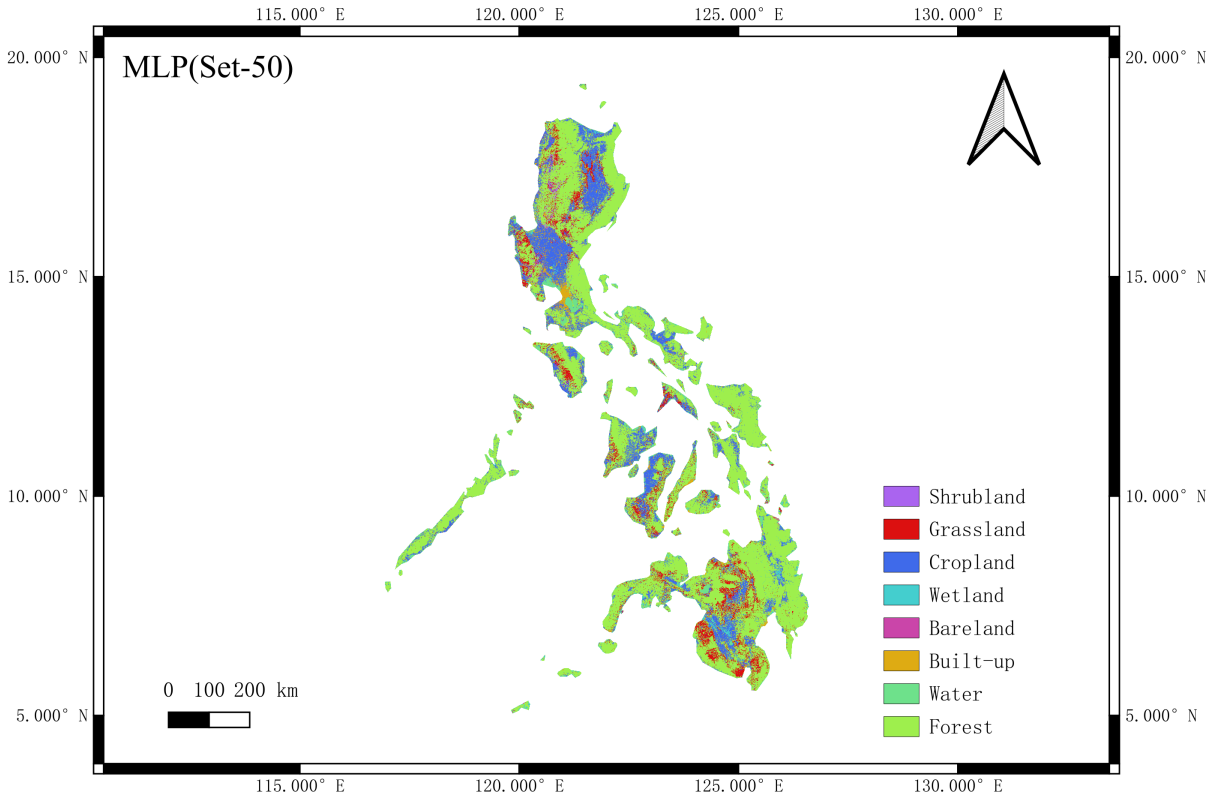


Figure 4.37: Land Cover Map by MLP with Set-50

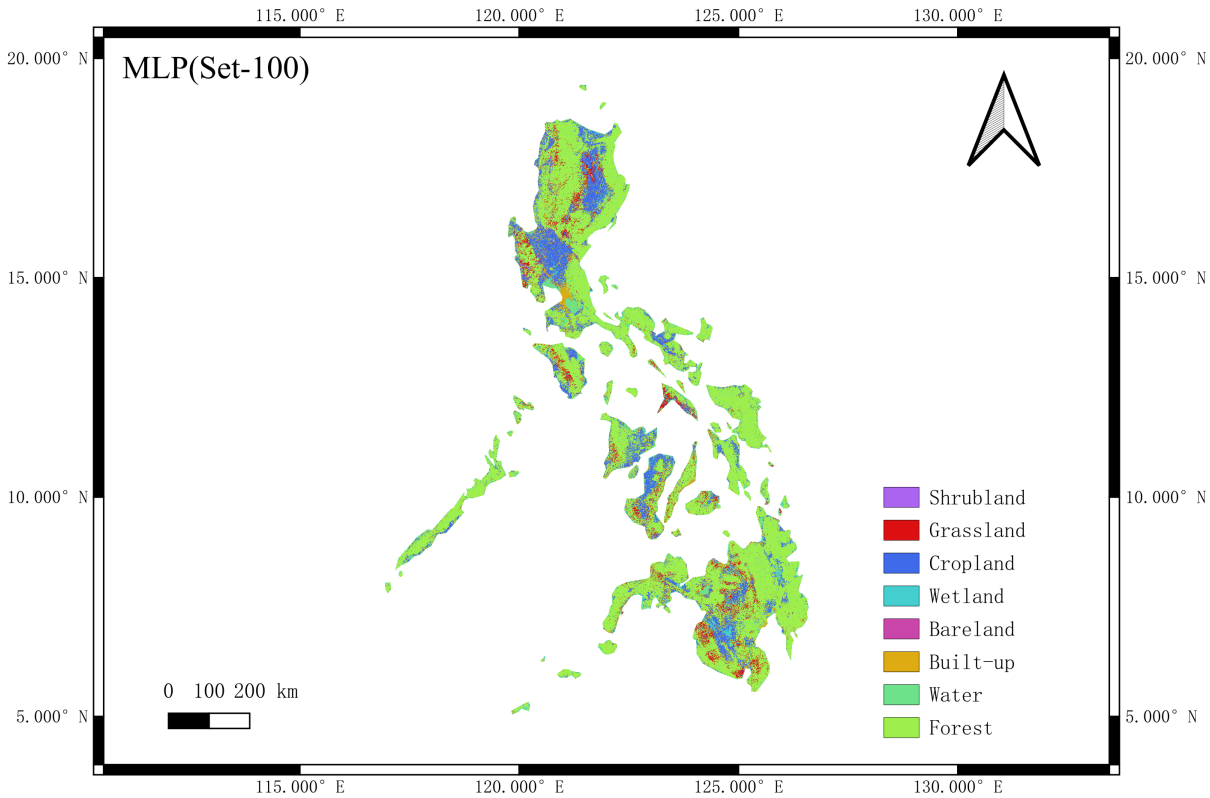


Figure 4.38: Land Cover Map by MLP with Set-100

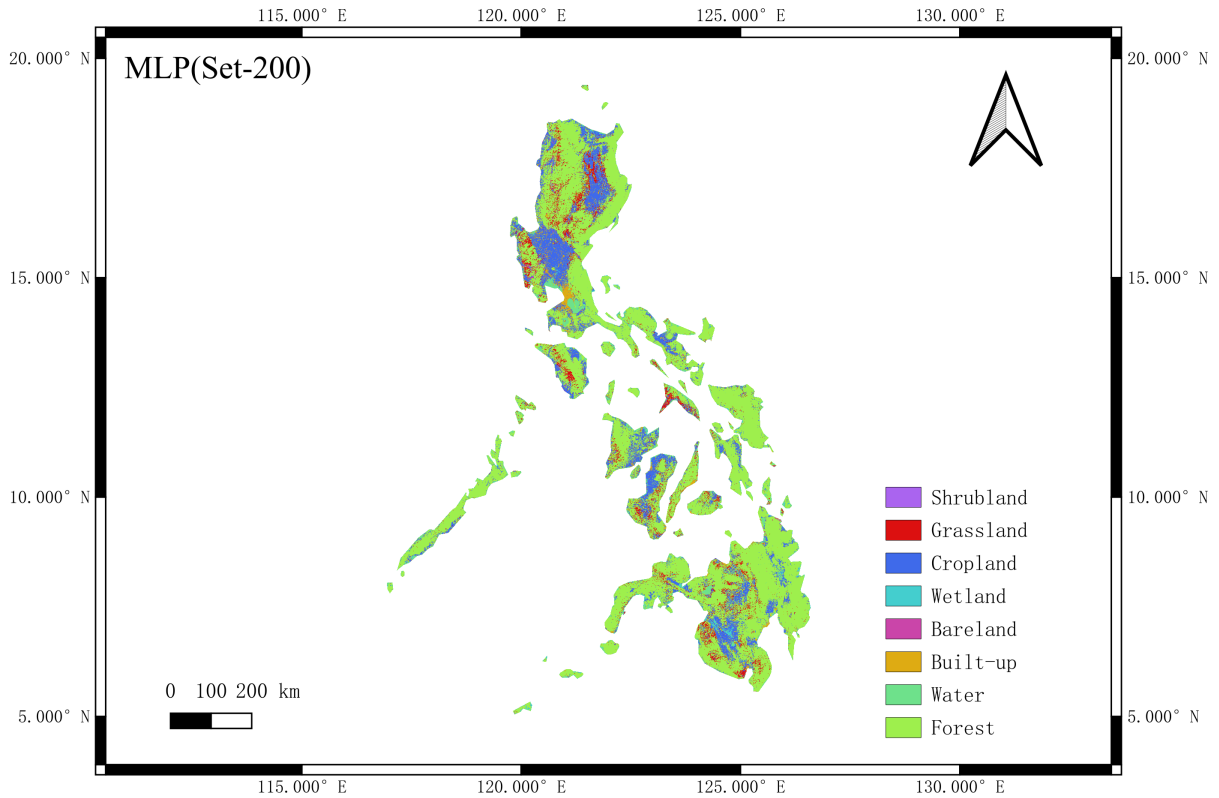


Figure 4.39: Land Cover Map by MLP with Set-200

4.3.2.3. Statistical Analysis

Tables 4.10 and 4.11 present the pixel counts and percentage distributions for the land cover maps generated by the RF, SVM, and MLP semi-supervised methods. Overall, they follow the same pattern with the existing multi-class HRLC dataset as illustrated in Section 2.7.

In general, the three models exhibit consistent trends regarding the dominant land cover classes. "Forest" constitutes the majority of the study area across all classifiers, ranging from roughly 67% to 71%. However, notable discrepancies are observed in the classification of specific minor classes. The "Grassland" coverage predicted by RF (3.64%) is distinctly lower than that of SVM and MLP (both approximately 6.5%). "Shrubland" and "Bareland" remain the least represented classes across all methods, each accounting for less than 1% of the total area.

The most significant divergence is observed in the "Wetland" class, where MLP-100B predicts a notably higher proportion (4.40%) compared to RF-50B (1.10%), SVM-50B (1.23%) and existing HRLC datasets. Manual inspection confirmed that this substantial overestimation stems from the misclassification of transition zones between vegetated ar-

eas and small and narrow water bodies (e.g., streams and small rivers). These transitional areas were misclassified as "Wetland" despite the fact that the training sets do not include these areas. Although this issue affects both MLP-50B and MLP-100B to varying degrees, it is effectively resolved in MLP-200B. The performance against SRS_val also reflects this issue.

Consequently, due to this significant overestimation, MLP-100B was excluded from further analysis. Instead, MLP-200B was selected, as it exhibits a class distribution consistent with other HRLC maps while maintaining an overall performance ranking second only to MLP-100B. This highlights that MLP needs more training samples to better generalize complex classes.

Class	MLP-50B	MLP-100B	MLP-200B	RF-50B	SVM-50B
Forest	1,919,723,899	2,005,431,124	2,053,703,137	2,142,732,529	2,055,999,415
Cropland	538,600,241	445,007,193	487,936,870	495,826,820	506,026,675
Grassland	244,245,021	196,549,382	197,776,229	108,902,455	195,846,711
Water	116,150,710	118,930,495	122,224,173	120,224,846	115,436,886
Built-up	80,222,328	78,081,668	75,153,108	61,985,143	64,709,421
Wetland	54,296,269	131,736,198	37,466,390	32,842,665	36,902,271
Bareland	17,417,571	13,076,539	16,409,059	14,742,315	15,404,727
Shrubland	22,125,664	3,969,104	2,112,737	15,524,930	2,455,597
Total	2,992,781,703	2,992,781,703	2,992,781,703	2,992,781,703	2,992,781,703

Table 4.10: Pixel Count Statistics for LC Maps by Semi-supervised Approach

Class	MLP-50B	MLP-100B	MLP-200B	RF-50B	SVM-50B
Forest	64.1451%	67.0089%	68.6219%	71.5967%	68.6986%
Cropland	17.9966%	14.8694%	16.3038%	16.5674%	16.9082%
Grassland	8.1611%	6.5674%	6.6084%	3.6388%	6.5440%
Water	3.8810%	3.9739%	4.0840%	4.0172%	3.8572%
Built-up	2.6805%	2.6090%	2.5111%	2.0712%	2.1622%
Wetland	1.8142%	4.4018%	1.2519%	1.0974%	1.2330%
Bareland	0.5820%	0.4369%	0.5483%	0.4926%	0.5147%
Shrubland	0.7393%	0.1326%	0.0706%	0.5187%	0.0821%

Table 4.11: Percentage Statistics for LC Maps by Semi-supervised Approach

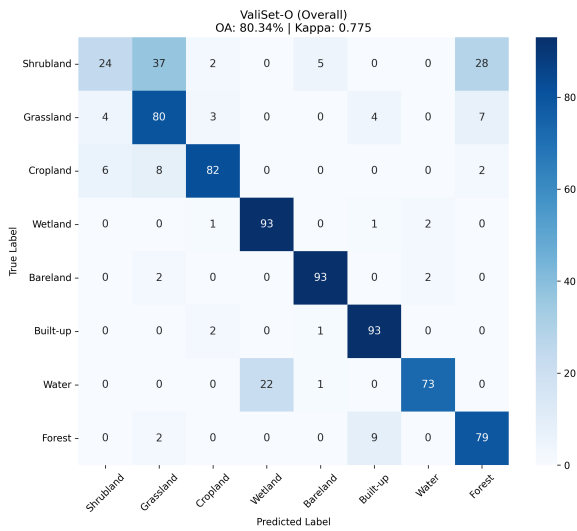
4.3.2.4. Accuracy Assessment

The following performance analysis is based on the 3 land cover maps produced with 3 classifiers with their optimal configuration (SVM-50B, RF-50B, MLP-200B). Both ValiSet-O and SRS_val are included in this analysis.

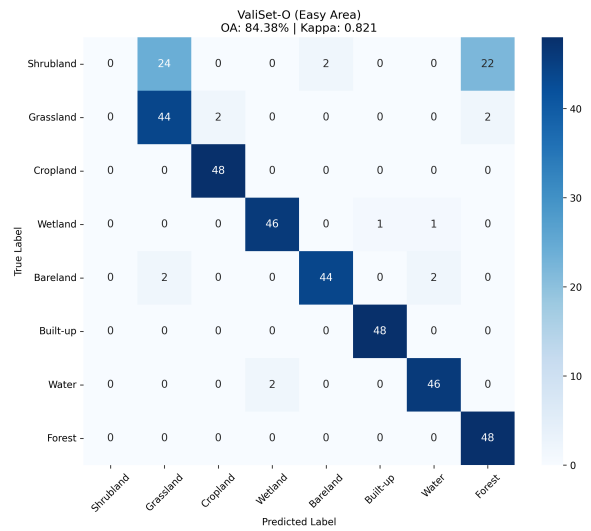
Figures 4.40, 4.41 and 4.42 show the confusion matrices on ValiSet-O (including its two subsets) and SRS_val for RF, SVM and MLP respectively.

As expected, all three classifiers (RF, SVM, and MLP) exhibit high accuracy on both validation sets in the study area. And all of them have a better performance on the easy area than on the hard area of ValiSet-O. However, no "Shrubland" is classified correctly in the easy area due to the limitation of ValiSet-O mentioned in Section 3.6. RF shows the largest gap between ValiSet-O (easy) and ValiSet-O (hard). Trees can easily overfit the clear rules of the "Easy" set but may struggle with the ambiguous rules. MLP has the best performance on the hard area of ValiSet-O, which means that it has excellent generalization ability.

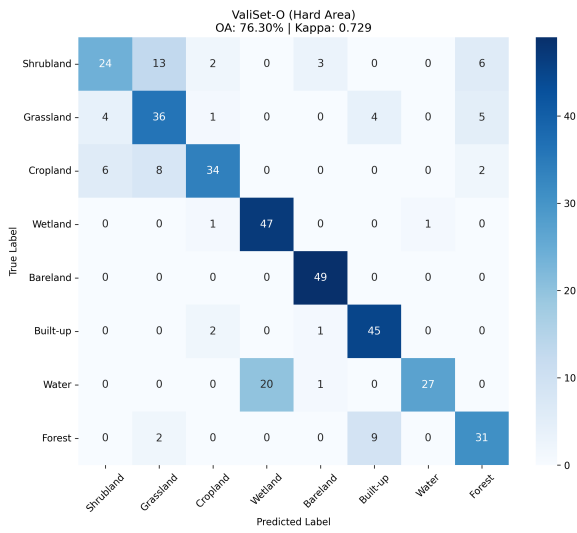
In general, the performance of every class is similar to the MOLCA and land cover maps derived from MOLCOPA, confirming the reliability of one another.



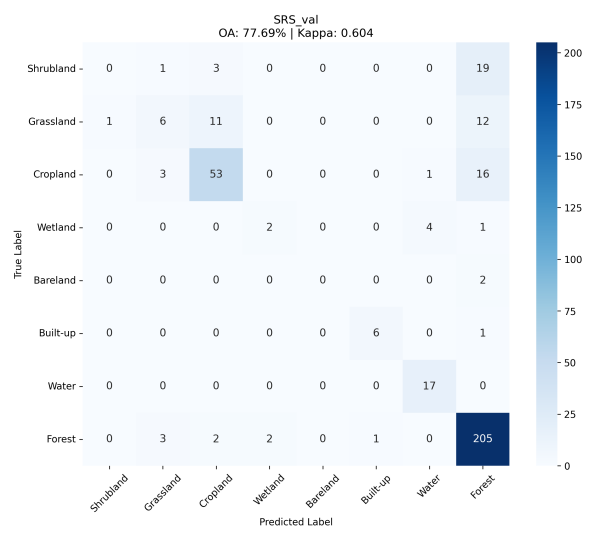
(a) ValiSet-O (All Samples)



(b) ValiSet-O (Easy Area)



(c) ValiSet-O (Hard Area)



(d) SRS_val

Figure 4.40: Confusion Matrices for RF

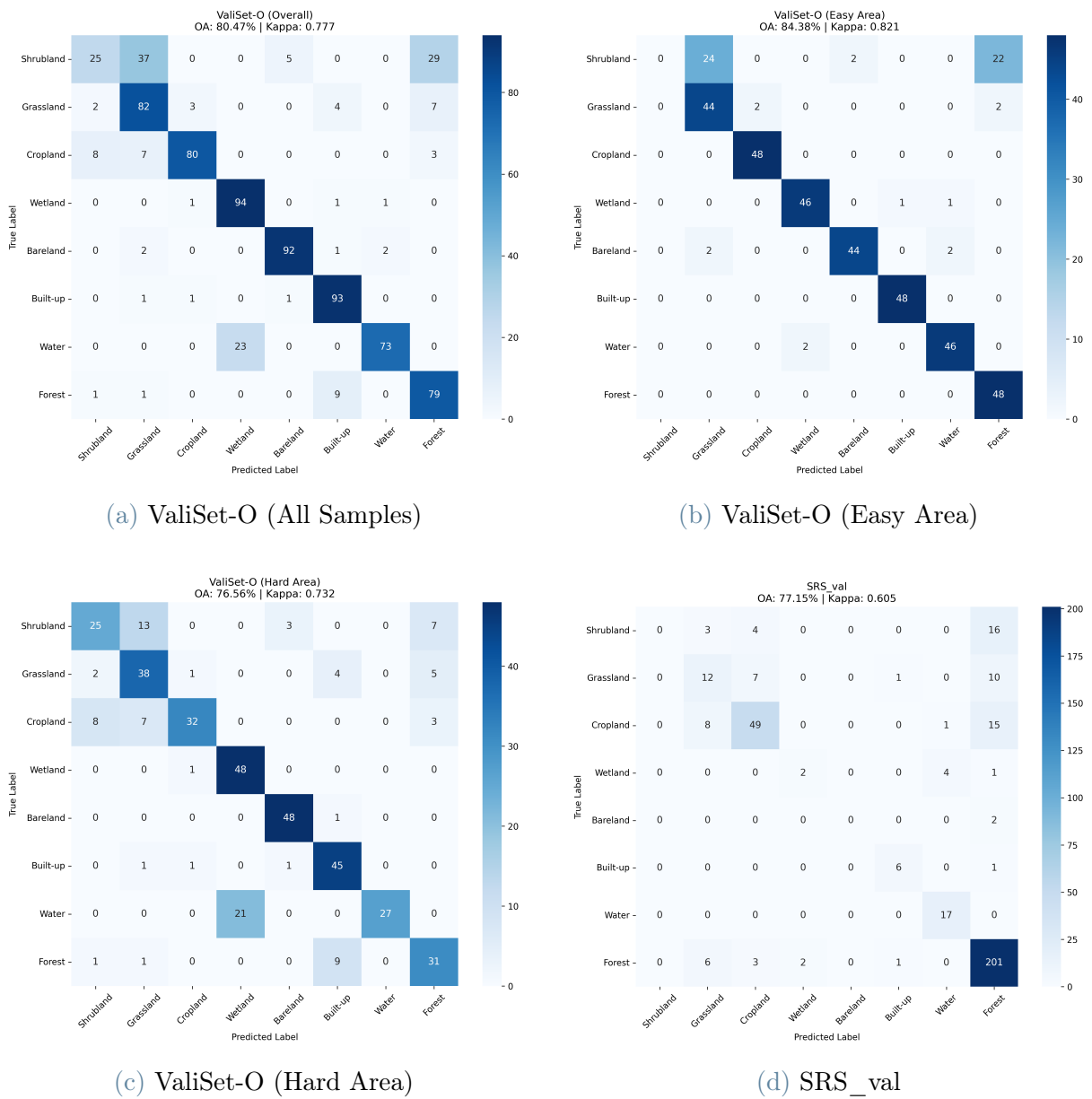
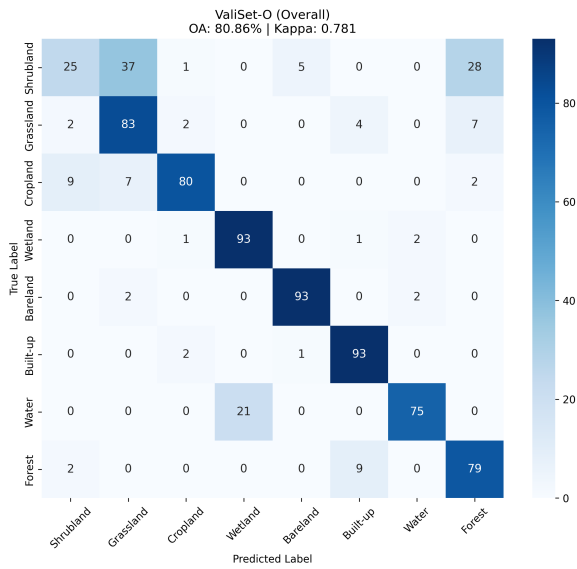
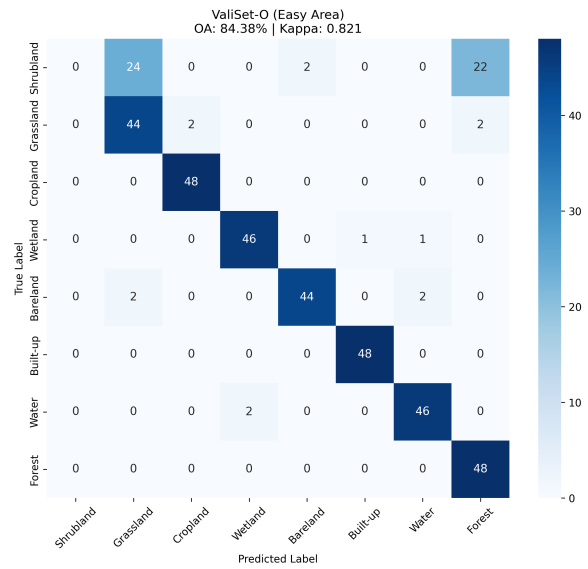


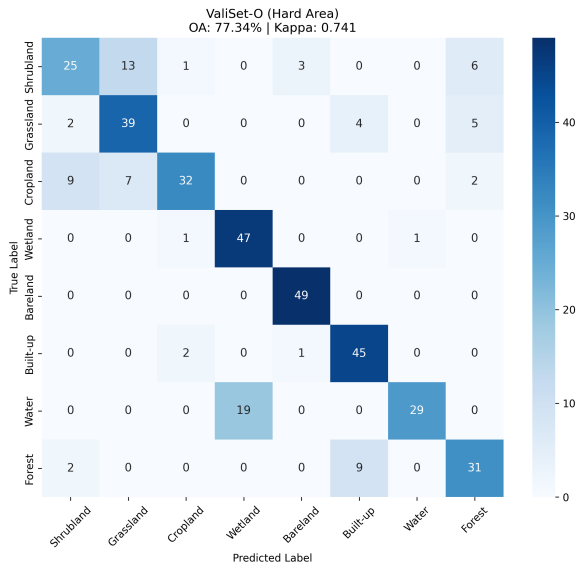
Figure 4.41: Confusion Matrices for SVM



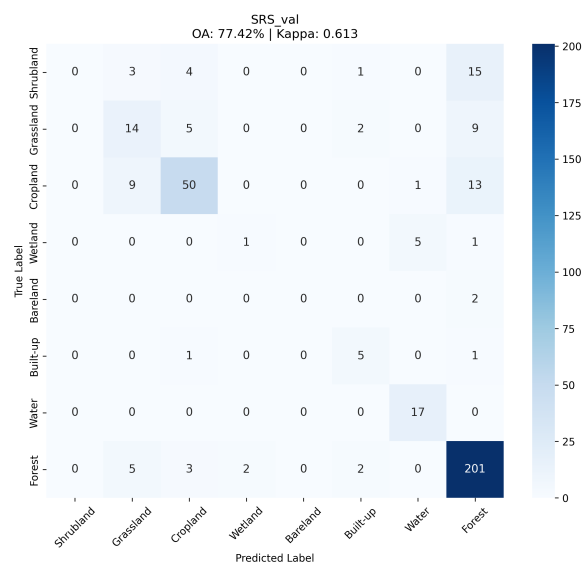
(a) ValiSet-O (All Samples)



(b) ValiSet-O (Easy Area)



(c) ValiSet-O (Hard Area)



(d) SRS_val

Figure 4.42: Confusion Matrices for MLP

4.3.3. Comprehensive Comparison

This section includes the comparison among abovementioned models based on ValiSet-O. Also, the comparison with 7 existing multi-class HRLC datasets based on both ValiSet-O and SRS_val.

4.3.3.1. Comparison among Models

Figure 4.43 shows the overall performance (Overall Accuracy and Kappa Coefficient) of different models on ValiSet-O. Overall, the semi-supervised approach always outperforms the unsupervised approach; all three supervised classifiers achieve an overall accuracy of around 80% and a Kappa Coefficient of 0.78. In contrast, the unsupervised approach achieves the best performance with an Overall Accuracy of 61.1% under local labeling (5-km window size) and the Kappa Coefficient is relatively much lower than Overall Accuracy caused by its poor ability in identifying minor classes.

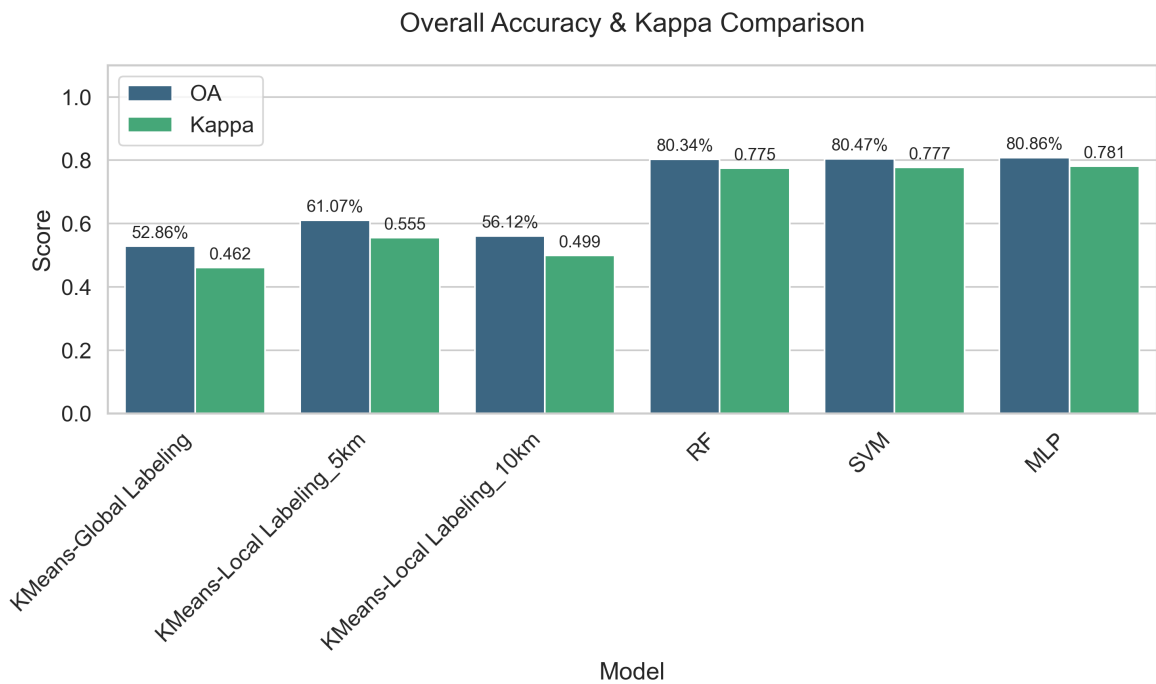


Figure 4.43: Comparison of Models in Overall Performance

Figures 4.44, 4.45, and 4.46 show the class-wise comparison (F1-score, Recall, and Precision) of different models.

The comparative analysis reveals a distinct performance among the evaluated methods, with supervised learning methods significantly outperforming the unsupervised baseline.

Three supervised classifiers show comparable performance. The MLP demonstrated the best generalization capability. It not only achieves the highest overall metrics but also maintains the most balanced Precision-Recall trade-off. Its primary advantage lies in its ability to mitigate misclassification in vegetation areas, outperforming other models in the difficult "Shrubland" and "Grassland" classes.

The primary bottleneck for model performance is the vegetation classes ("Forest," "Grassland," and "Shrubland"). Classes such as "Water," "Built-up," and "Bareland" are identified with high accuracy across all supervised models ($F1 > 0.90$ for MLP and RF). "Shrubland" is the most challenging category, with F1-scores remaining below 0.40 for all models. "Grassland" and "Forest" also show moderate confusion.

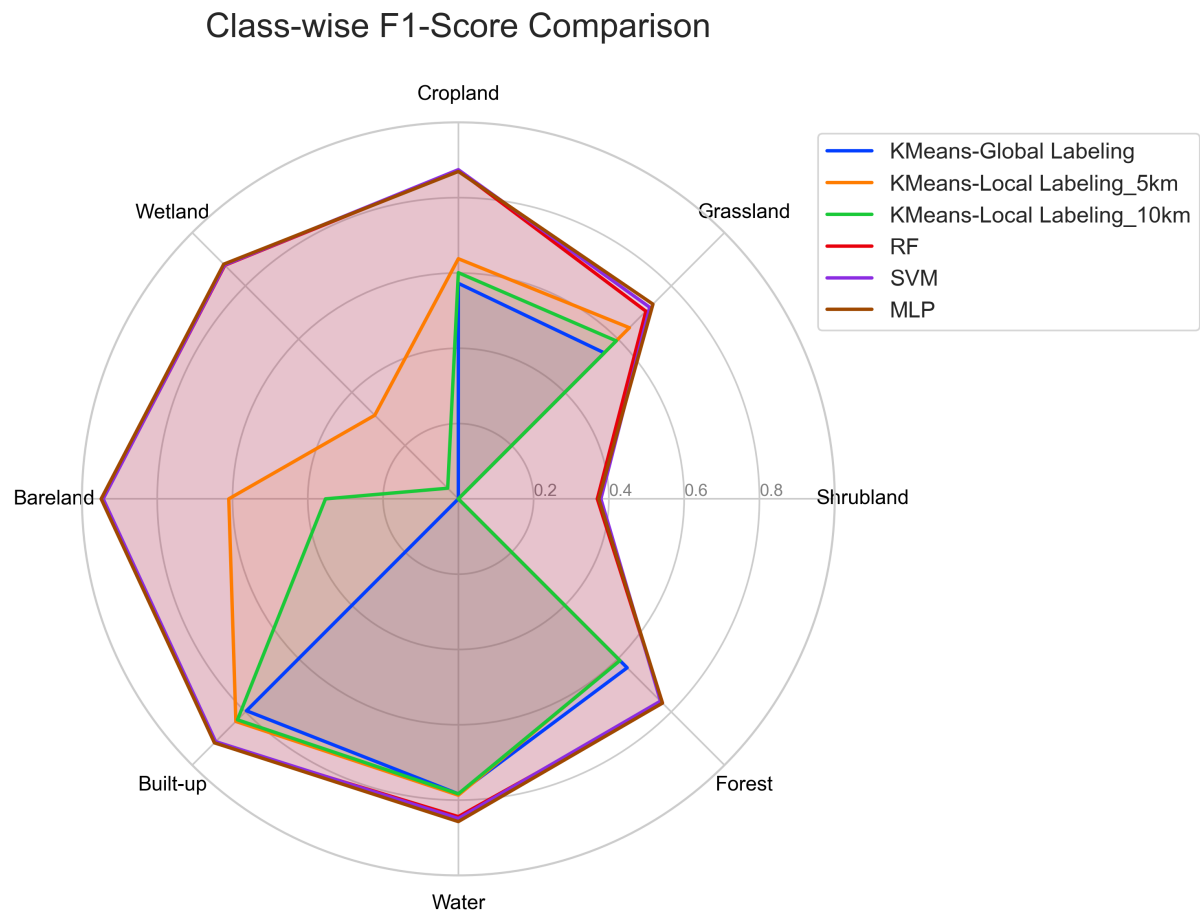


Figure 4.44: Comparison of Models in F1-Score



Figure 4.45: Comparison of Models in Recall

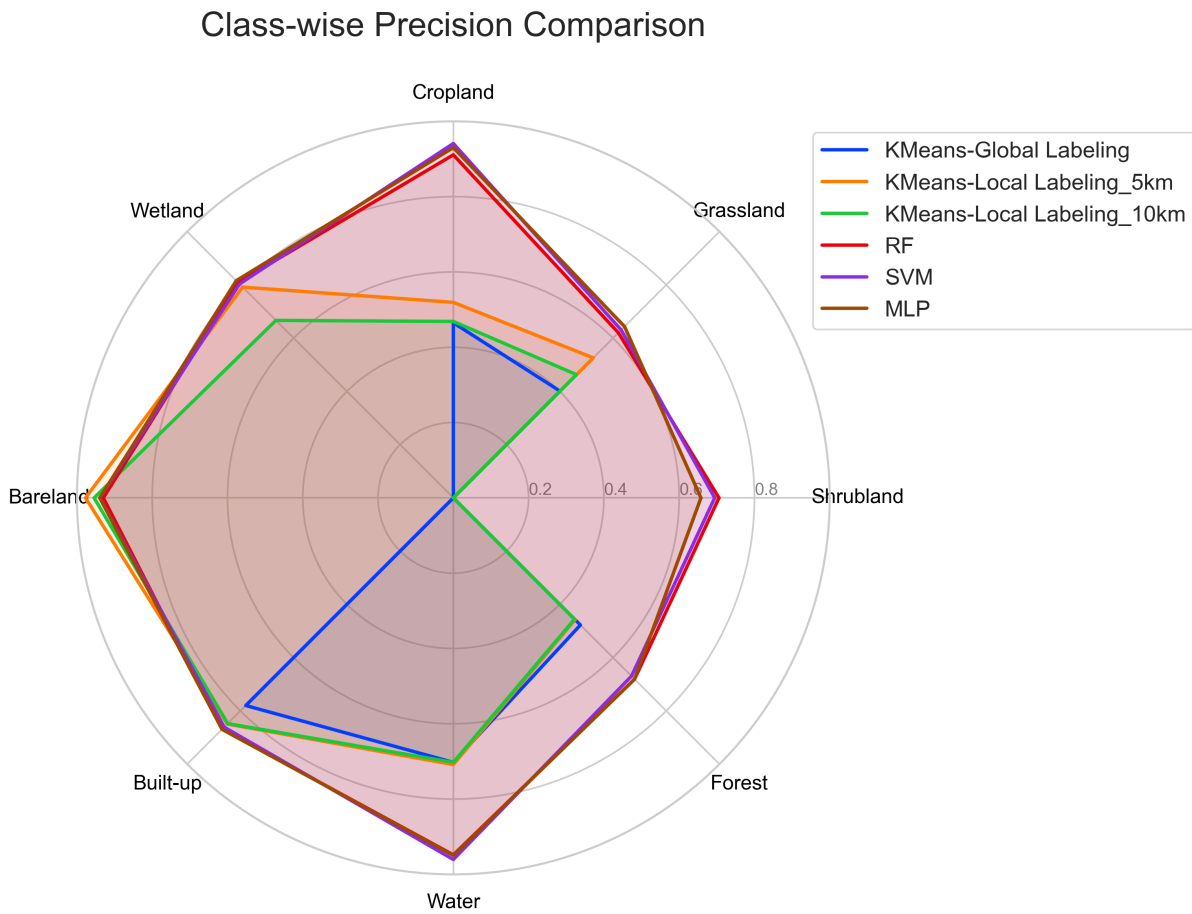


Figure 4.46: Comparison of Models in Precision

4.3.3.2. Comparison with Existing Datasets

Since MLP outperforms all other models in this study, it was selected to make a comparison with existing HRLC datasets. Figures 4.47, 4.48, and 4.49 illustrate the performance of existing datasets and MLP. The support varies across datasets because some classes of existing HRLC datasets cannot be mapped to MOLCA code or it presents "NoData." Some classes present an F1-Score of 0 because those classes cannot be mapped to MOLCA code except for GLC-FCS30D. As illustrated in Section 2.7, GLC-FCS30D has only 1138 pixels classified as "Grassland," probably resulting from its finer classification system.

Among the existing datasets, Dynamic World, ESRI LULC, and WorldCover exhibit the strongest performance, with Overall Accuracy hovering around 80%. The remaining datasets all fall below 70%. This also explains why MOLCA 2010 and 2015 performed significantly worse than the 2019 version: they lack the integration of these high-accuracy datasets.

Overall, the performance of the MLP is excellent, especially considering it was trained using approximately 1400 sample points. In terms of Overall Accuracy and the Kappa Coefficient, the MLP trails only ESRI LULC. However, since the ESRI LULC is not mapped to the difficult-to-classify "Shrubland" and "Grassland" classes, its overall accuracy may be overestimated. Furthermore, the MLP's performance across classes shows a high degree of consistency with that of other products. Classes that are inherently difficult to distinguish are challenging for all products, while easily classifiable classes consistently yield higher accuracy. In terms of individual class performance, the MLP is on par with other existing products.

This reveals the feasibility of extracting training samples ("silver points") from MOLCA and the validity of MOLCOPA.

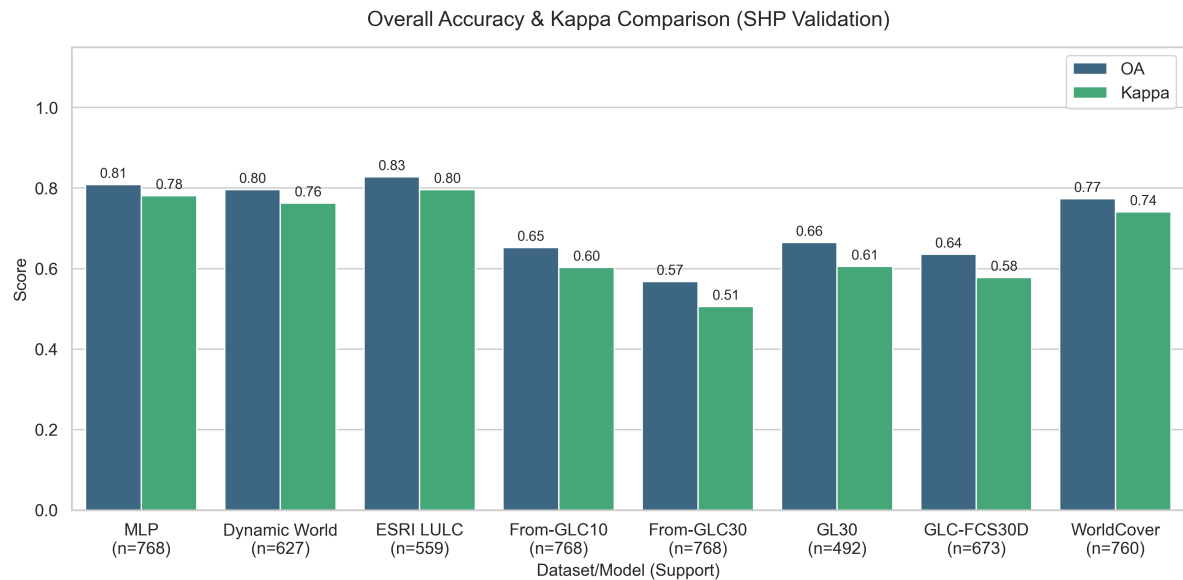


Figure 4.47: Comparison with Existing Datasets in Overall Performance

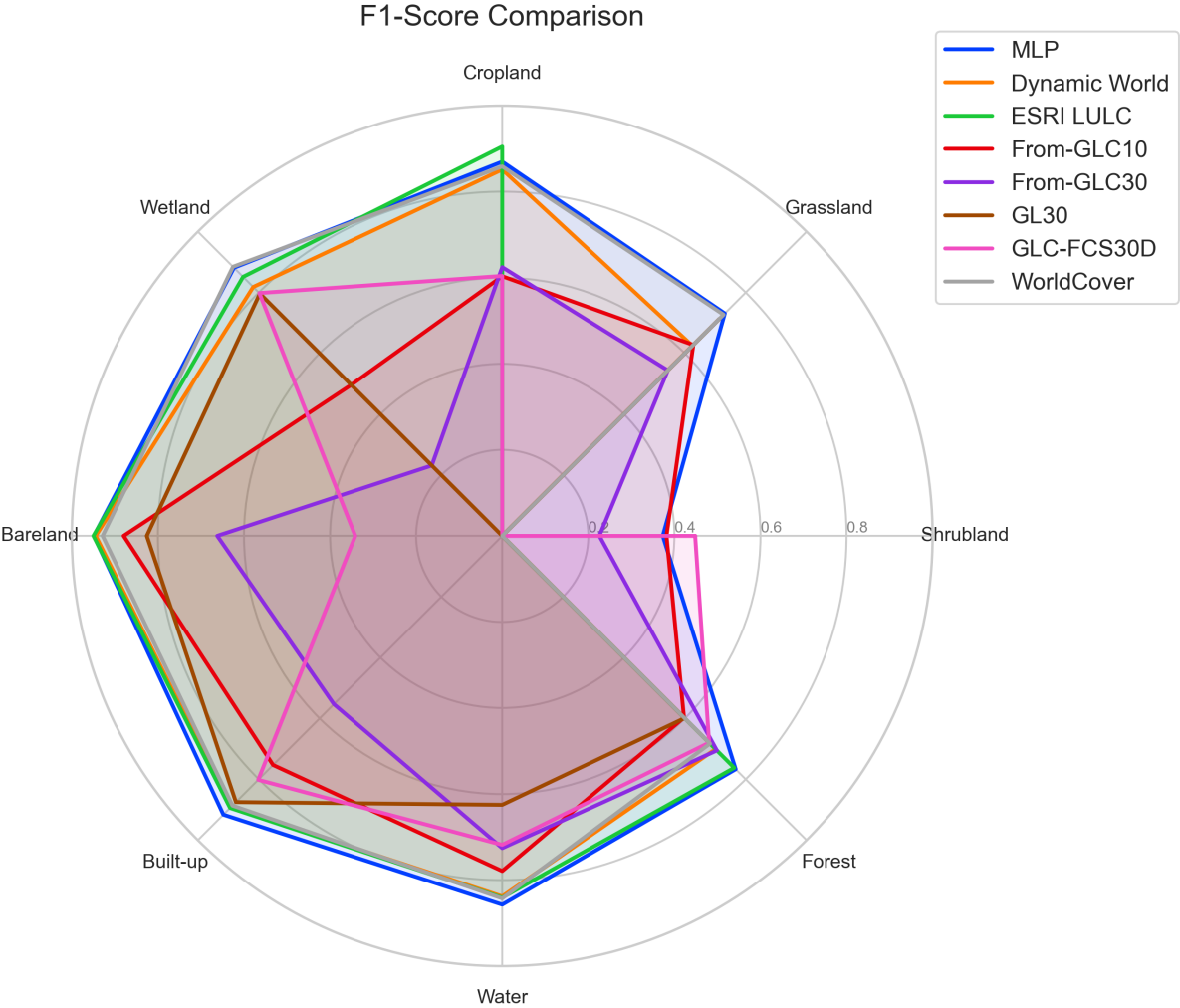


Figure 4.48: Comparison with Existing Datasets in F1-Score

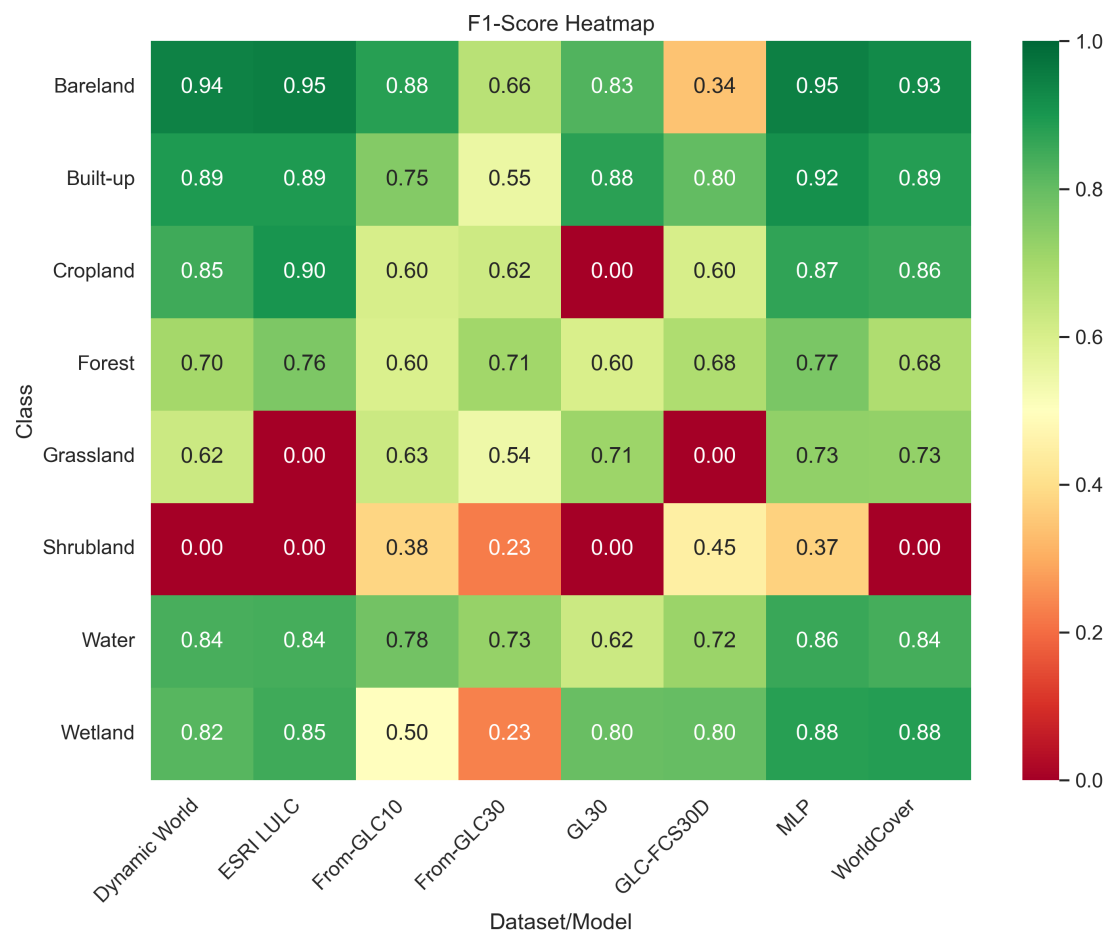


Figure 4.49: Comparison with Existing Datasets in F1-Score (Detailed)

5 | Conclusion

This study established a comprehensive framework for HRLC mapping, starting from generating a benchmark land cover map (MOLCA) to probabilistic partial agreement methods (MOLCOPA) , and finally to GeoAI-based land cover mapping with the benchmark land cover map and probabilistic products.

MOLCA was proven accurate enough to be a benchmark land cover map. Even though the quality of MOLCA is intrinsically linked to the quantity and quality of input multi-class HRLC datasets. MOLCA 2019 achieved the highest performance (Overall Accuracy = 92.71%) due to the integration of 14 input datasets. However, the unanimous consensus requirement resulted in significant data gaps, with "NoData" covering approximately 40% of the study area, which highlighted the limitation of unanimous consensus.

MOLCOPA overcame coverage limitations by treating land cover datasets as probability. It was highlighted that a clear trade-off between coverage and reliability. Strict strategies (like Major Agreement, $\tau = 0.9$) yielded high accuracy, but they resulted in significant "NoData" gaps similar to MOLCA. In addition, balanced approaches such as "N-1 Agreement" and "Hybrid Agreement" successfully fill the "NoData" gaps without sacrificing much quality, demonstrating the effectiveness of this probabilistic consensus approach. With MOLCOPA and the count map, customized land cover maps can be produced easily to satisfy various requirements.

Integration of GeoAI algorithms with the generated products (MOLCA and MOLCOPA) was performed to generate a complete land cover map.

Unsupervised and Semi-supervised were tried utilizing the products in different ways. Unsupervised approaches proved to be limited, showing a strong bias toward majority classes under various strategies. In addition, unsupervised approaches are likely to produce a incomplete land cover maps (with a spatial coverage $> 99\%$) due to gaps in MOLCOPA.

In contrast, the semi-supervised framework demonstrated superior performance; all three classifiers applied in this study show strong performance. MLP emerged as the most robust classifier, achieving an Overall Accuracy of 81% and a Kappa Coefficient of 0.78 with only

1407 training samples. Comparing with other existing HRLC datasets, the trained MLP model is better or comparable to them from the perspective of overall performance and individual class performance. Additionally, the participation of MOLCOPA significantly improved the performance of models. These reveal high data efficiency of the MOLCA and validity of MOLCOPA.

For future work, other MOLCOPA-derived land cover maps are also potentially to be the sources of training samples. They might provide more information in the hard-to-classify area, resulting in better generalization ability of the models. Interpolation can be considered as an alternative to fill the gaps in MOLCOPA, resulting fully complete land maps using unsupervised approaches and better results for other approaches. Additionally, MOLCA and MOLCOPA are also possibly useful in Deep Learning.

Bibliography

- [1] B. Bauer-Marschallinger and K. Falkner. Wasting petabytes: A survey of the Sentinel-2 UTM tiling grid and its spatial overhead. *ISPRS Journal of Photogrammetry and Remote Sensing*, 202:682–690, Aug. 2023. ISSN 0924-2716. doi: 10.1016/j.isprsjprs.2023.07.015.
- [2] A. Bey, A. Sánchez-Paus Díaz, D. Maniatis, G. Marchi, D. Mollicone, S. Ricci, J.-F. Bastin, R. Moore, S. Federici, M. Rezende, C. Patriarca, R. Turia, G. Gamoga, H. Abe, E. Kaidong, and G. Miceli. Collect Earth: Land Use and Land Cover Assessment through Augmented Visual Interpretation. *Remote Sensing*, 8(10):807, Oct. 2016. ISSN 2072-4292. doi: 10.3390/rs8100807.
- [3] G. Bratic, D. Oxoli, and M. A. Brovelli. Map of Land Cover Agreement: Ensambling Existing Datasets for Large-Scale Training Data Provision. *Remote Sensing*, 15(15):3774, jul 2023. ISSN 2072-4292. doi: 10.3390/rs15153774.
- [4] G. Bratic, V. Yordanov, and M. A. Brovelli. High-resolution land cover classification: cost-effective approach for extraction of reliable training data from existing land cover datasets. *International Journal of Digital Earth*, 16(1):3618–3636, Oct. 2023. ISSN 1753-8947. doi: 10.1080/17538947.2023.2253784.
- [5] L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, Oct. 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324.
- [6] C. F. Brown, S. P. Brumby, B. Guzder-Williams, T. Birch, S. B. Hyde, J. Mazzariello, W. Czerwinski, V. J. Pasquarella, R. Haertel, S. Ilyushchenko, K. Schwehr, M. Weisse, F. Stolle, C. Hanson, O. Guinan, R. Moore, and A. M. Tait. Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data*, 9(1):251, June 2022. ISSN 2052-4463. doi: 10.1038/s41597-022-01307-4.
- [7] C. F. Brown, M. R. Kazmierski, V. J. Pasquarella, W. J. Rucklidge, M. Samsikova, C. Zhang, E. Shelhamer, E. Lahera, O. Wiles, S. Ilyushchenko, N. Gorelick, L. L. Zhang, S. Alj, E. Schechter, S. Askay, O. Guinan, R. Moore, A. Boukouvalas, and

- P. Kohli. AlphaEarth foundations: An embedding field model for accurate and efficient global mapping from sparse label data, 2025. doi: 10.48550/arXiv.2507.22291.
- [8] Central Intelligence Agency. Philippines - The World Factbook, n.d. URL <https://www.cia.gov/the-world-factbook/countries/philippines/>. (Last Accessed on January 27, 2026).
- [9] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16: 321–357, June 2002. ISSN 1076-9757. doi: 10.1613/jair.953.
- [10] J. Chen, J. Chen, A. Liao, X. Cao, L. Chen, X. Chen, C. He, G. Han, S. Peng, M. Lu, W. Zhang, X. Tong, and J. Mills. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103:7–27, May 2015. ISSN 09242716. doi: 10.1016/j.isprsjprs.2014.09.002.
- [11] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, editors, *Computer Vision – ECCV 2018*, pages 833–851, Cham, 2018. Springer International Publishing. ISBN 978-3-030-01234-2. doi: 10.1007/978-3-030-01234-2_49.
- [12] W. G. Cochran. *Sampling Techniques*. John Wiley & Sons, 1977.
- [13] C. Corbane, M. Pesaresi, P. Politis, V. Syrris, A. J. Florczyk, P. Soille, L. Maffenini, A. Burger, V. Vasilev, D. Rodriguez, F. Sabo, L. Dijkstra, and T. Kemper. Big earth data analytics on Sentinel-1 and Landsat imagery in support to global human settlements mapping. *Big Earth Data*, 1(1-2):118–144, Dec. 2017. ISSN 2096-4471, 2574-5417. doi: 10.1080/20964471.2017.1397899.
- [14] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, Sept. 1995. ISSN 1573-0565. doi: 10.1007/BF00994018.
- [15] A. V. Dorogush, V. Ershov, and A. Gulin. CatBoost: gradient boosting with categorical features support, Oct. 2018. doi: 10.48550/arXiv.1810.11363.
- [16] T. Esch, M. Marconcini, A. Felbier, A. Roth, W. Heldens, M. Huber, M. Schwinger, H. Taubenböck, A. Müller, and S. Dech. Urban Footprint Processor—Fully Automated Processing Chain Generating Settlement Masks from Global Data of the TanDEM-X Mission. *IEEE Geoscience and Remote Sensing Letters*, 10(6):1617–1621, 2013. doi: 10.1109/LGRS.2013.2272953.
- [17] European Commission. Joint Research Centre. *Mapping global forest cover of the year*

- 2020 to support the EU regulation on deforestation-free supply chains. Publications Office, LU, 2024. doi: 10.2760/262532.
- [18] European Space Agency (ESA). SENTINEL-2 User Handbook, 2015. URL https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook. (Last Accessed on January 27, 2026).
- [19] FAO. LAND COVER CLASSIFICATION SYSTEM, n.d. URL <https://www.fao.org/4/x0596e/x0596e01e.htm>. (Last Accessed on February 17, 2026).
- [20] J. J. Feddema, K. W. Oleson, G. B. Bonan, L. O. Mearns, L. E. Buja, G. A. Meehl, and W. M. Washington. The Importance of Land-Cover Change in Simulating Future Climates. *Science*, 310(5754):1674–1678, Dec. 2005. doi: 10.1126/science.1118160.
- [21] Z. Feng, C. Atzberger, S. Jaffer, J. Knezevic, S. Sormunen, R. Young, M. C. Lisaius, M. Immitzer, T. Jackson, J. Ball, D. A. Coomes, A. Madhavapeddy, A. Blake, and S. Keshav. TESSERA: Temporal Embeddings of Surface Spectra for Earth Representation and Analysis, Nov. 2025. doi: 10.48550/arXiv.2506.20380.
- [22] M. Ferraguti, S. Magallanes, M. Suarez-Rubio, P. J. J. Bates, A. Marzal, and S. C. Renner. Does land-use and land cover affect vector-borne diseases? A systematic review and meta-analysis. *Landscape Ecology*, 38(10):2433–2451, Oct. 2023. ISSN 1572-9761. doi: 10.1007/s10980-023-01746-3.
- [23] P. Gong, J. Wang, L. Yu, Y. Zhao, Y. Zhao, L. Liang, Z. Niu, X. Huang, H. Fu, S. Liu, C. Li, X. Li, W. Fu, C. Liu, Y. Xu, X. Wang, Q. Cheng, L. Hu, W. Yao, H. Zhang, P. Zhu, Z. Zhao, H. Zhang, Y. Zheng, L. Ji, Y. Zhang, H. Chen, A. Yan, J. Guo, L. Yu, L. Wang, X. Liu, T. Shi, M. Zhu, Y. Chen, G. Yang, P. Tang, B. Xu, C. Giri, N. Clinton, Z. Zhu, J. Chen, and J. Chen. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, 34(7):2607–2654, Apr. 2013. ISSN 0143-1161, 1366-5901. doi: 10.1080/01431161.2012.748992.
- [24] P. Gong, H. Liu, M. Zhang, C. Li, J. Wang, H. Huang, N. Clinton, L. Ji, W. Li, Y. Bai, B. Chen, B. Xu, Z. Zhu, C. Yuan, H. Ping Suen, J. Guo, N. Xu, W. Li, Y. Zhao, J. Yang, C. Yu, X. Wang, H. Fu, L. Yu, I. Dronova, F. Hui, X. Cheng, X. Shi, F. Xiao, Q. Liu, and L. Song. 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. *Science Bulletin*, 64(6):370–373, Mar. 2019. ISSN 20959273. doi: 10.1016/j.scib.2019.03.002.
- [25] Google. Satellite Embedding V1, n.d. URL <https://developers.google.com/>

earth-engine/datasets/catalog/GOOGLE_SATELLITE_EMBEDDING_V1_ANNUAL.

(Last Accessed on February 17, 2026).

- [26] J. W. Hager, L. L. Fry, S. S. Jacks, and D. R. Hill. Datums, Ellipsoids, Grids, and Grid Reference Systems:. Technical report, Defense Technical Information Center, Fort Belvoir, VA, Jan. 1992.
- [27] J. Han, Z. Zhang, Y. Luo, J. Cao, L. Zhang, F. Cheng, H. Zhuang, J. Zhang, and F. Tao. NESEA-Rice10: high-resolution annual paddy rice maps for Northeast and Southeast Asia from 2017 to 2019. *Earth System Science Data*, 13(12):5969–5986, 2021. doi: 10.5194/essd-13-5969-2021.
- [28] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342(6160):850–853, 2013. doi: 10.1126/science.1244693.
- [29] Japan Aerospace Exploration Agency (JAXA). *Dataset Description: PALSAR-2 Global Forest/Non-Forest (FNF) (Version 2.0.0)*. Earth Observation Research Center (EORC), Japan, 2018. URL https://www.eorc.jaxa.jp/ALOS/en/dataset/pdf/DatasetDescription_PALSAR2_FNF_v200a.pdf. (Last Accessed on January 02, 2026).
- [30] K. Karra, C. Kontgis, Z. Statman-Weil, J. C. Mazzariello, M. Mathis, and S. P. Brumby. Global land use / land cover with Sentinel 2 and deep learning. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 4704–4707, Brussels, Belgium, July 2021. IEEE. ISBN 978-1-6654-0369-6. doi: 10.1109/IGARSS47720.2021.9553499.
- [31] Karthik and B. R. Shivakumar. Land Cover Mapping Capability of Chaincluster, K-Means, and ISODATA techniques—A Case Study. In S. Kalya, M. Kulkarni, and K. S. Shivaprakasha, editors, *Advances in VLSI, Signal Processing, Power Electronics, IoT, Communication and Embedded Systems*, pages 273–288, Singapore, 2021. Springer Singapore. ISBN 978-981-16-0443-0.
- [32] Z. Lv, T. Liu, C. Shi, J. A. Benediktsson, and H. Du. Novel Land Cover Change Detection Method Based on k-Means Clustering and Adaptive Majority Voting Using Bitemporal Remote Sensing Images. *IEEE Access*, 7:34425–34437, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2892648.
- [33] J. MacQueen. Some methods for classification and analysis of multivariate observa-

- tions. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California press Oakland, CA, USA, 1967.
- [34] M. Marconcini, A. Metz-Marconcini, S. Üreyen, D. Palacios-Lopez, W. Hanke, F. Bachofer, J. Zeidler, T. Esch, N. Gorelick, A. Kakarla, M. Paganini, and E. Strano. Outlining where humans live, the World Settlement Footprint 2015. *Scientific Data*, 7(1):242, July 2020. ISSN 2052-4463. doi: 10.1038/s41597-020-00580-5.
- [35] M. Marconcini, A. Metz-Marconcini, T. Esch, and N. Gorelick. Understanding Current Trends in Global Urbanisation - The World Settlement Footprint Suite. *GI_Forum*, 1:33–38, 2021. ISSN 2308-1708, 2308-1708. doi: 10.1553/giscience2021_01_s33.
- [36] National Aeronautics and Space Administration (NASA). Tiling System – Harmonized Landsat Sentinel-2, n.d. URL <https://hls.gsfc.nasa.gov/products-description/tiling-system/>. (Last Accessed on January 27, 2026).
- [37] Open Foris Team. Open Foris Collect, n.d. URL <https://www.openforis.org/>. (Last Accessed on February 17, 2026).
- [38] J.-F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633):418–422, Dec. 2016. ISSN 1476-4687. doi: 10.1038/nature20584.
- [39] M. Pesaresi, V. Syrris, and A. Julea. A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning. *Remote Sensing*, 8(5):399, May 2016. ISSN 2072-4292. doi: 10.3390/rs8050399.
- [40] M. Pesaresi, M. Schiavina, P. Politis, S. Freire, K. Krasnodebska, J. H. Uhl, A. Carioli, C. Corbane, L. Dijkstra, P. Florio, H. K. Friedrich, J. Gao, S. Leyk, L. Lu, L. Maffenini, I. Mari-Rivero, M. Melchiorri, V. Syrris, J. Van Den Hoek, and T. Kemper. Advances on the Global Human Settlement Layer by joint assessment of Earth Observation and population survey data. *International Journal of Digital Earth*, 17(1):2390454, Dec. 2024. ISSN 1753-8947, 1753-8955. doi: 10.1080/17538947.2024.2390454.
- [41] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4. doi: 10.1007/978-3-319-24574-4_28.

- [42] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958. ISSN 1939-1471. doi: 10.1037/h0042519.
- [43] A. Shadmehri Toosi, O. Batelaan, M. Shanafeld, and H. Guan. Land Use-Land Cover and Hydrological Modeling: A Review. *WIREs Water*, 12(2):e70013, 2025. ISSN 2049-1948. doi: 10.1002/wat2.70013.
- [44] E. Shelhamer, J. Long, and T. Darrell. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, 2017. doi: 10.1109/TPAMI.2016.2572683.
- [45] M. Shimada, T. Itoh, T. Motooka, M. Watanabe, T. Shiraishi, R. Thapa, and R. Lucas. New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment*, 155:13–31, 2014. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2014.04.014>.
- [46] G. Tseng, R. Cartuyvels, I. Zvonkov, M. Purohit, D. Rolnick, and H. Kerner. Lightweight, Pre-trained Transformers for Remote Sensing Timeseries, Feb. 2024. doi: 10.48550/arXiv.2304.14065.
- [47] A. Vali, S. Comai, and M. Matteucci. Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data: A Review. *Remote Sensing*, 12(15):2495, Jan. 2020. ISSN 2072-4292. doi: 10.3390/rs12152495.
- [48] R. Yu, W. Yu, and X. Wang. KAN or MLP: A Fairer Comparison, 2024. doi: 10.48550/ARXIV.2407.16674.
- [49] D. Zanaga, R. Van De Kerchove, W. De Keersmaecker, N. Souverijns, C. Brockmann, R. Quast, J. Wevers, A. Grosu, A. Paccini, S. Vergnaud, O. Cartus, M. Santoro, S. Fritz, I. Georgieva, M. Lesiv, S. Carter, M. Herold, L. Li, N.-E. Tsendbazar, F. Ramoino, and O. Arino. ESA WorldCover 10 m 2020 v100, Oct. 2021. doi: 10.5281/ZENODO.5571936.
- [50] D. Zanaga, R. Van De Kerchove, D. Daems, W. De Keersmaecker, C. Brockmann, G. Kirches, J. Wevers, O. Cartus, M. Santoro, S. Fritz, M. Lesiv, M. Herold, N.-E. Tsendbazar, P. Xu, F. Ramoino, and O. Arino. ESA WorldCover 10 m 2021 v200, Oct. 2022. doi: 10.5281/ZENODO.7254221.
- [51] X. Zhang, L. Liu, T. Zhao, Y. Gao, X. Chen, and J. Mi. GISD30: global 30 m impervious-surface dynamic dataset from 1985 to 2020 using time-series Landsat

- imagery on the Google Earth Engine platform. *Earth System Science Data*, 14(4): 1831–1856, 2022. doi: 10.5194/essd-14-1831-2022.
- [52] X. Zhang, L. Liu, T. Zhao, J. Wang, W. Liu, and X. Chen. Global annual wetland dataset at 30 m with a fine classification system from 2000 to 2022. *Scientific Data*, 11(1):310, Mar. 2024. ISSN 2052-4463. doi: 10.1038/s41597-024-03143-0.
- [53] X. Zhang, T. Zhao, H. Xu, W. Liu, J. Wang, X. Chen, and L. Liu. GLC_fcs30d: the first global 30 m land-cover dynamics monitoring product with a fine classification system for the period from 1985 to 2022 generated using dense-time-series Landsat imagery and the continuous change-detection method. *Earth System Science Data*, 16(3):1353–1381, Mar. 2024. ISSN 1866-3516. doi: 10.5194/essd-16-1353-2024.
- [54] T. Zhao, X. Zhang, Y. Gao, J. Mi, W. Liu, J. Wang, M. Jiang, and L. Liu. Assessing the Accuracy and Consistency of Six Fine-Resolution Global Land Cover Products Using a Novel Stratified Random Sampling Validation Dataset. *Remote Sensing*, 15(9):2285, 2023. ISSN 2072-4292. doi: 10.3390/rs15092285.
- [55] Z. Zhu and C. E. Woodcock. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment*, 144:152–171, Mar. 2014. ISSN 0034-4257. doi: 10.1016/j.rse.2014.01.011.

A | Appendix A

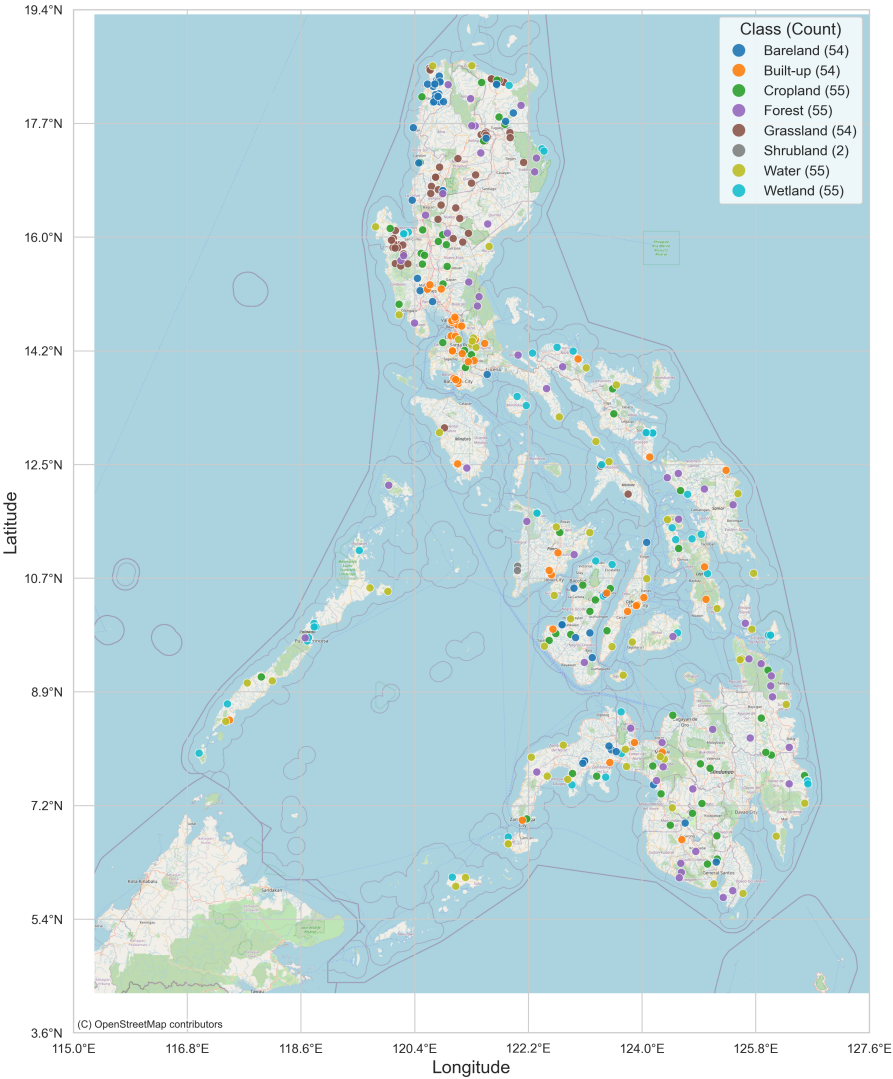


Figure A.1: Spatial Distribution of the Visual Interpretation Samples for MOLCA 2010

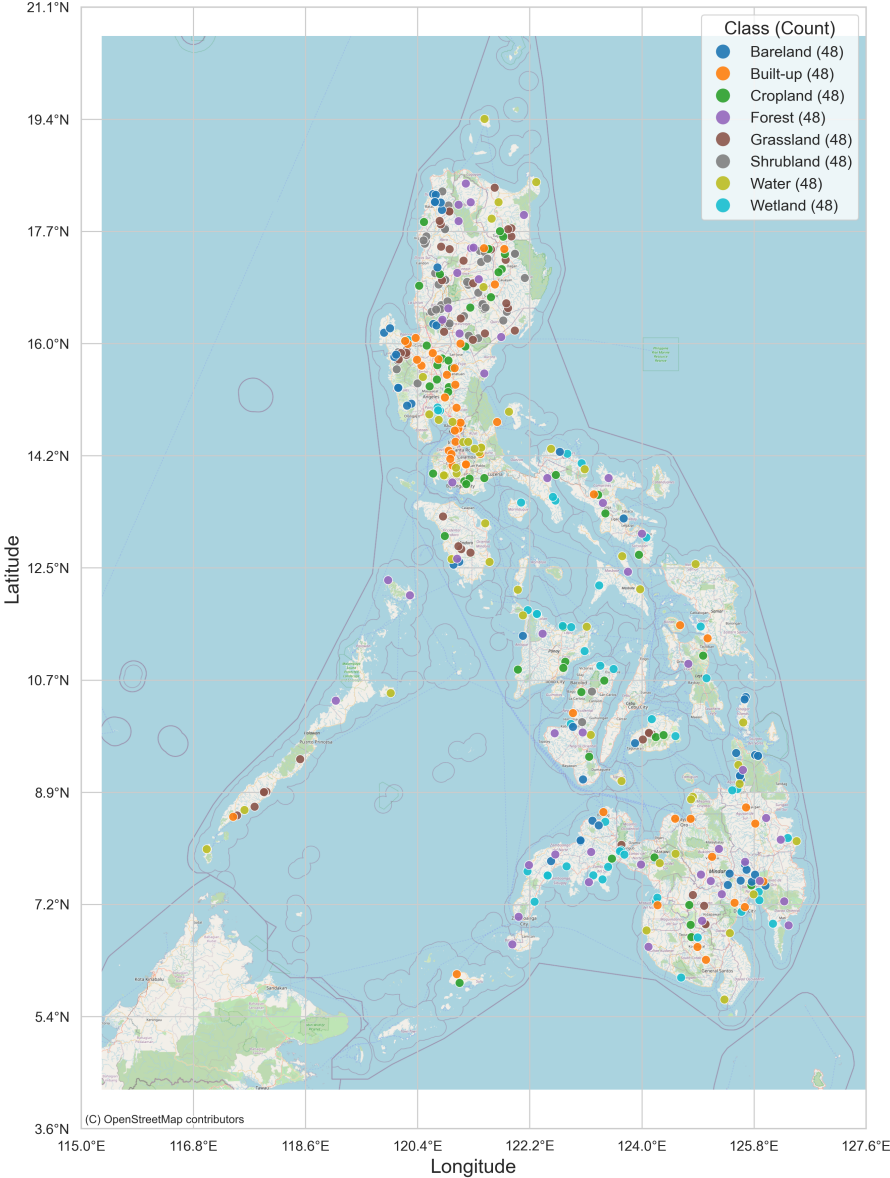


Figure A.2: Spatial Distribution of the Visual Interpretation Samples for MOLCA 2015

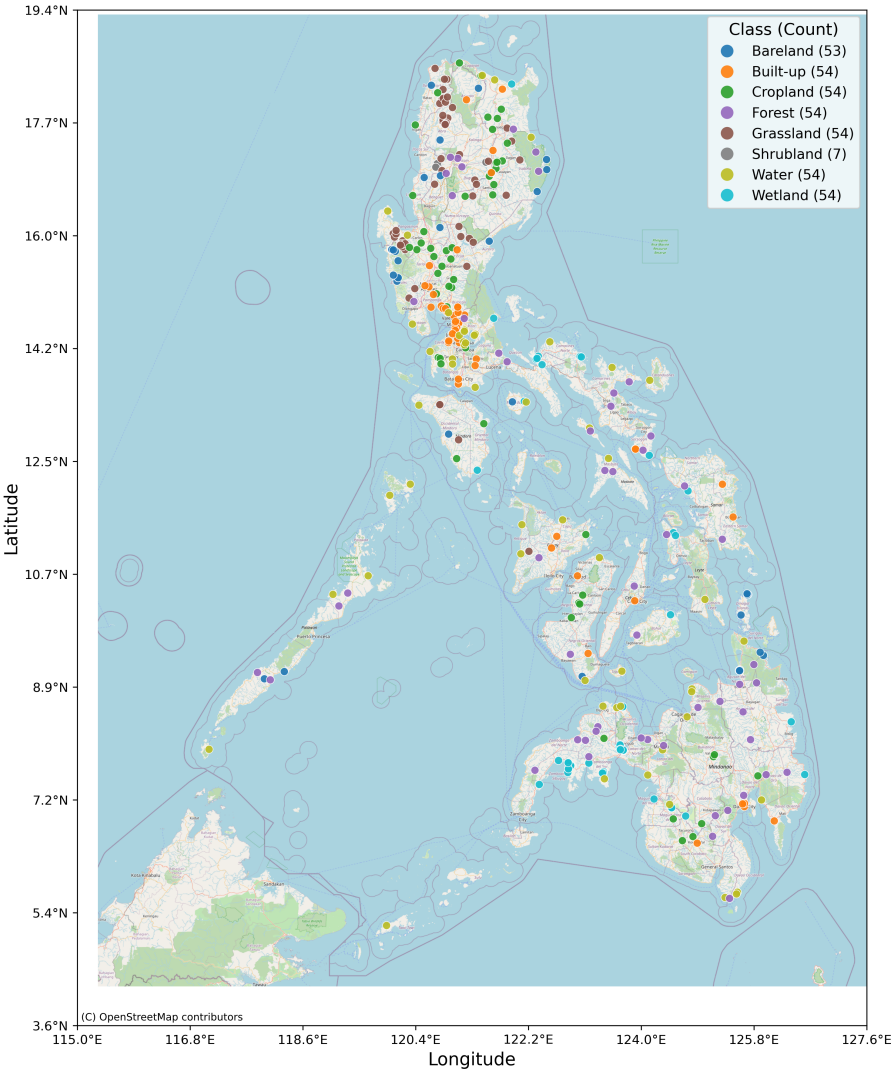


Figure A.3: Spatial Distribution of the Visual Interpretation Samples for MOLCA 2019 (Overall Agreement)

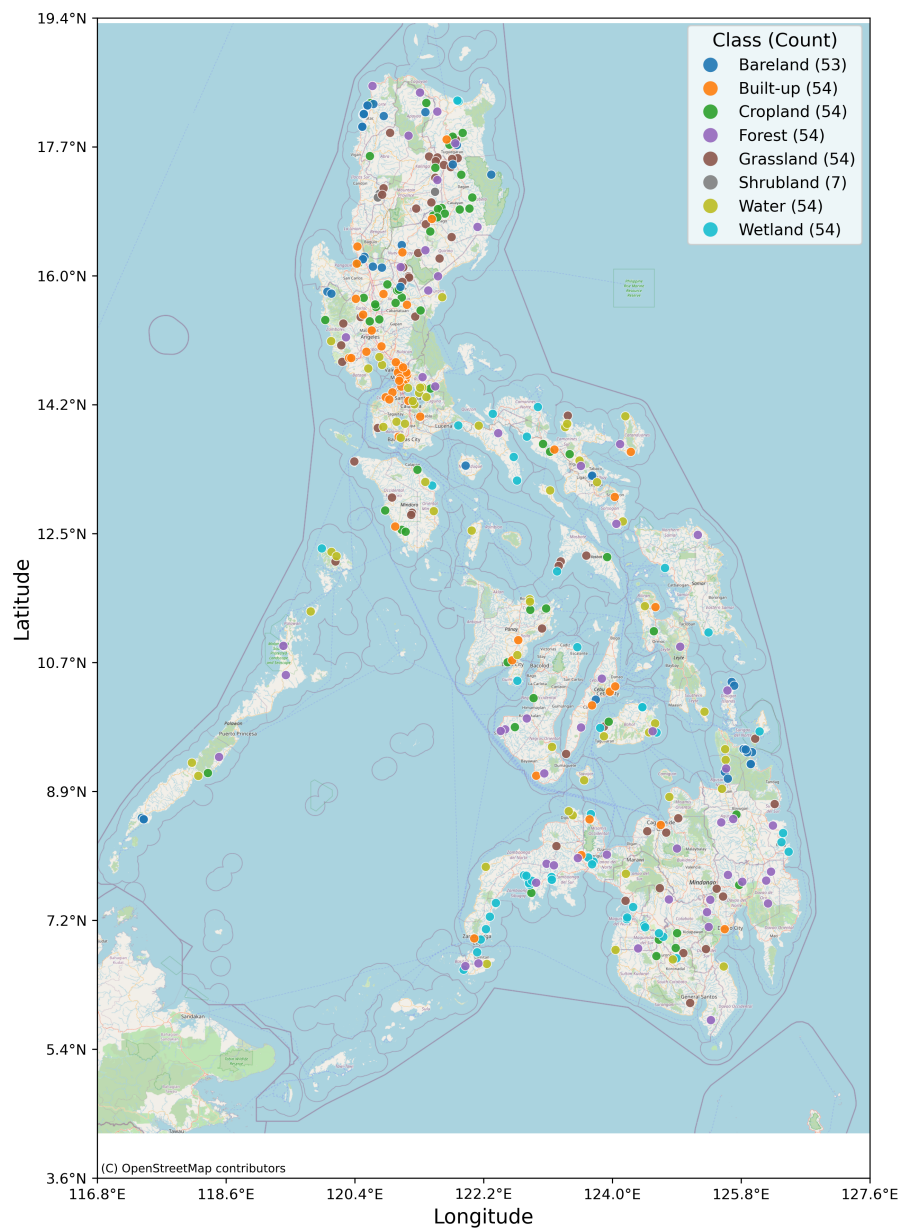


Figure A.4: Spatial Distribution of the Visual Interpretation Samples for Hybrid Agreement

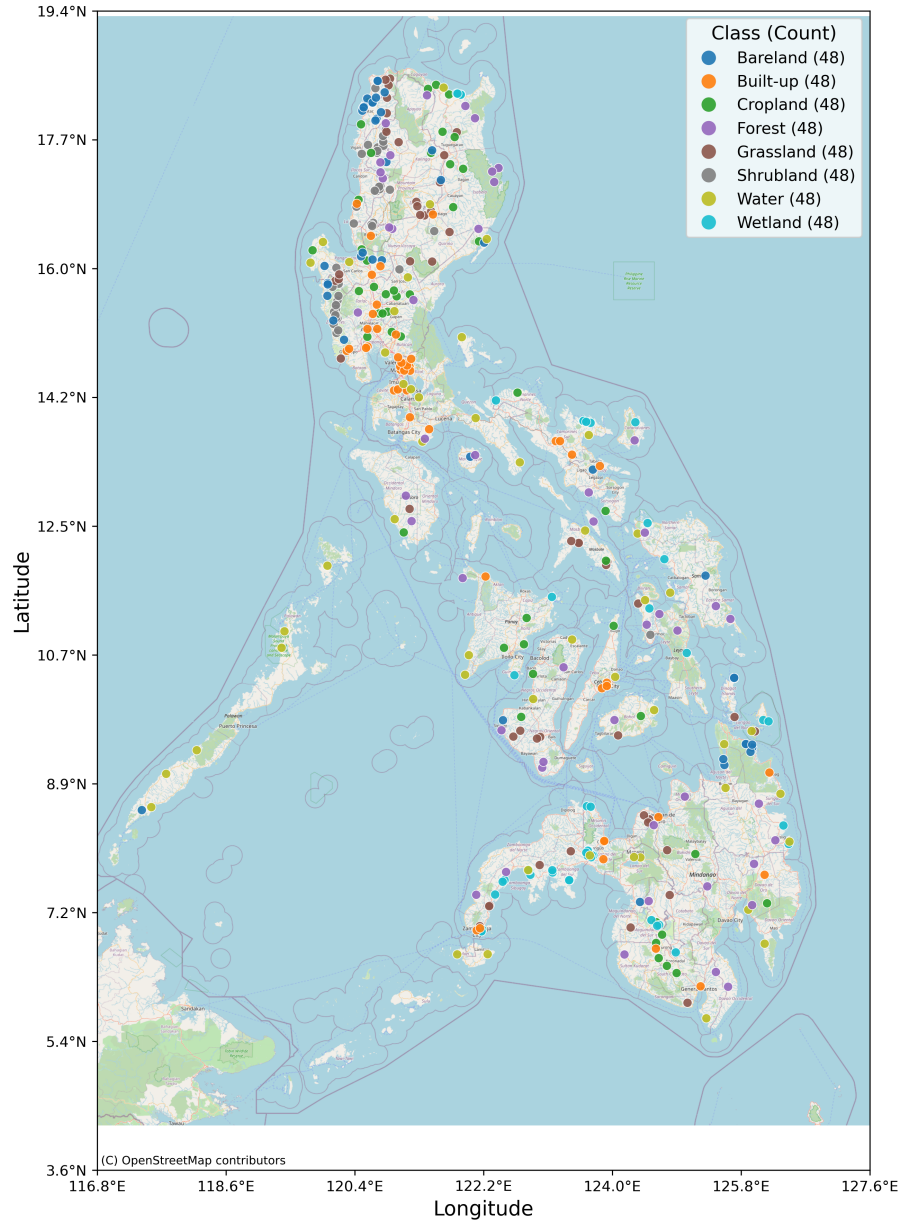


Figure A.5: Spatial Distribution of the Visual Interpretation Samples for N-1 Agreement

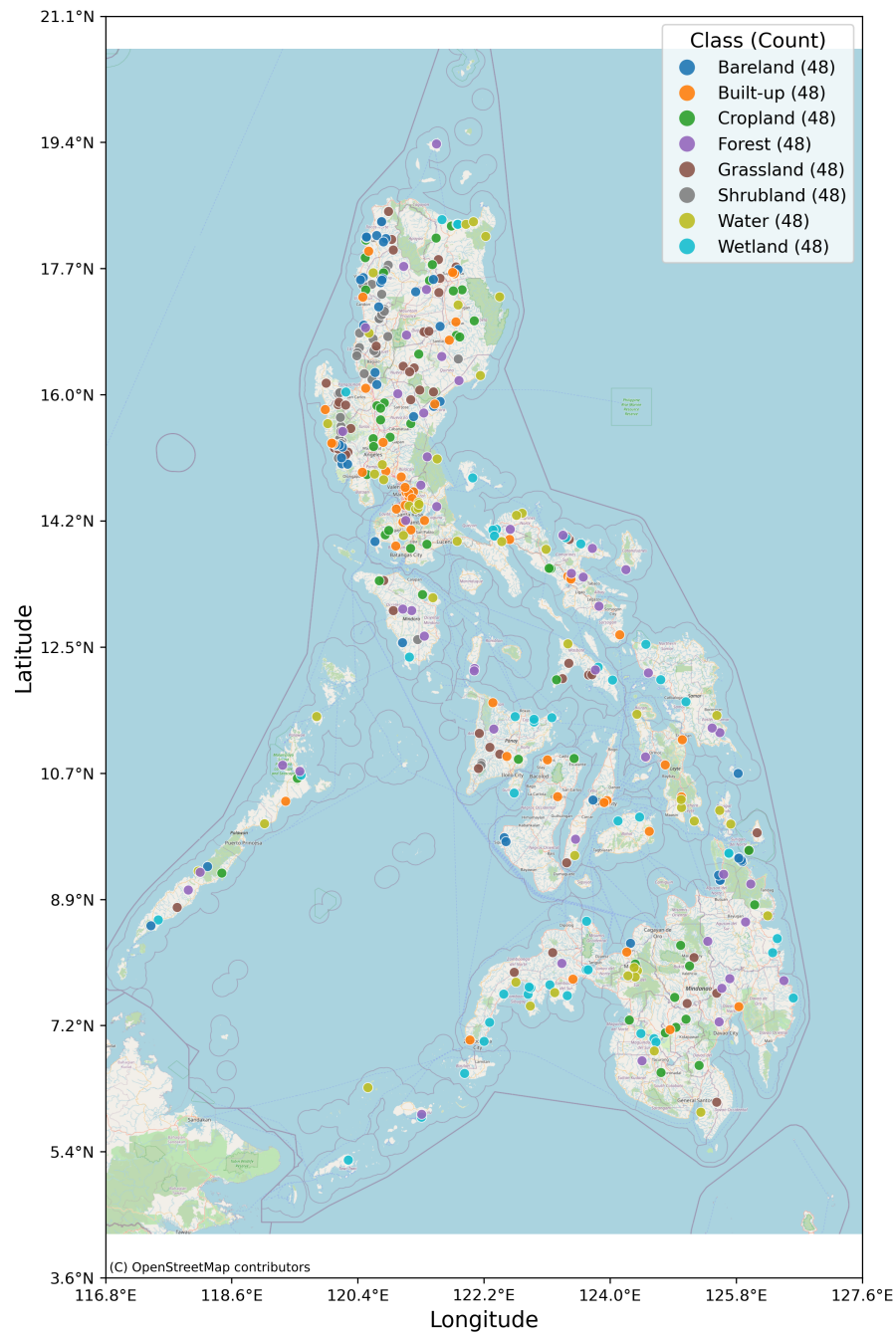


Figure A.6: Spatial Distribution of the Visual Interpretation Samples for Major Agreement ($\tau = 0.5$)

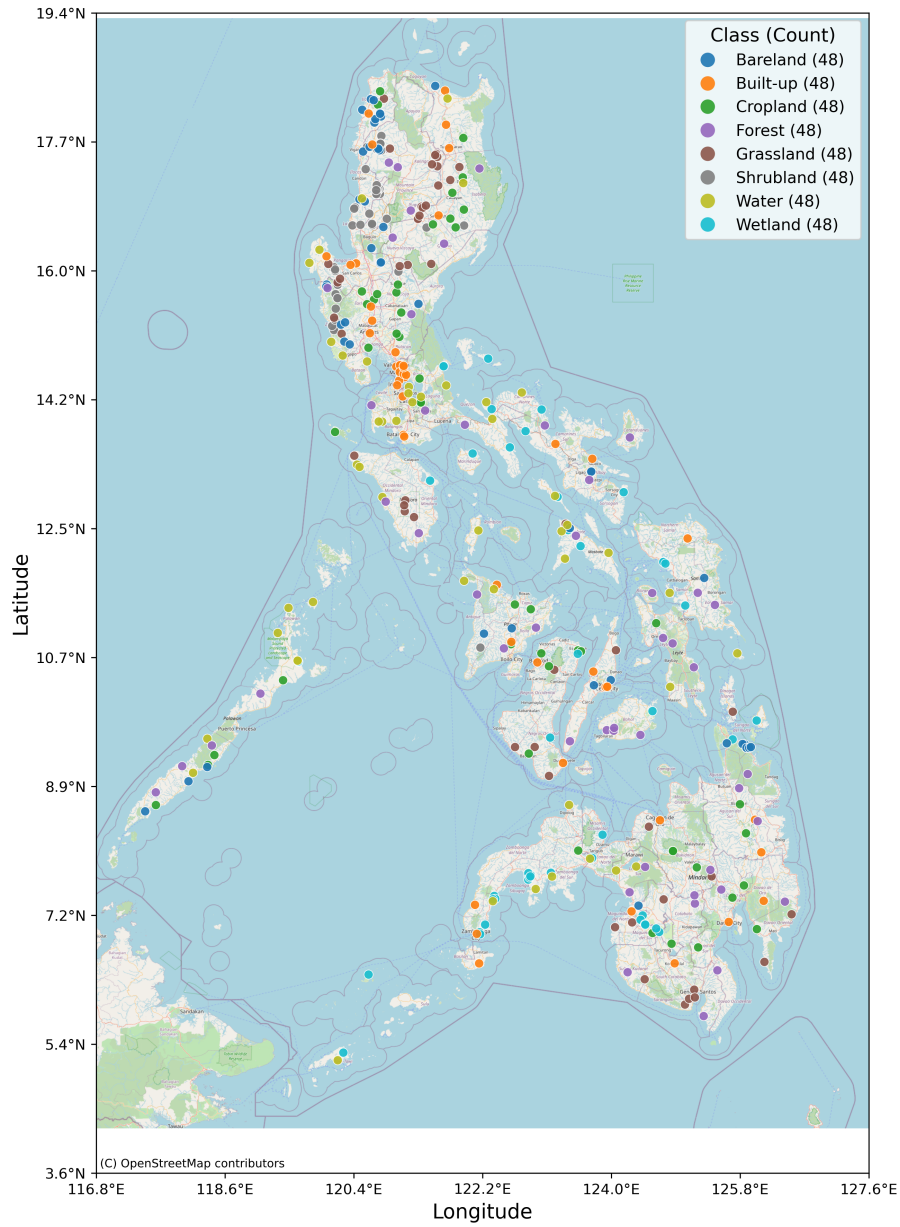


Figure A.7: Spatial Distribution of the Visual Interpretation Samples for Major Agreement ($\tau = 0.6$)

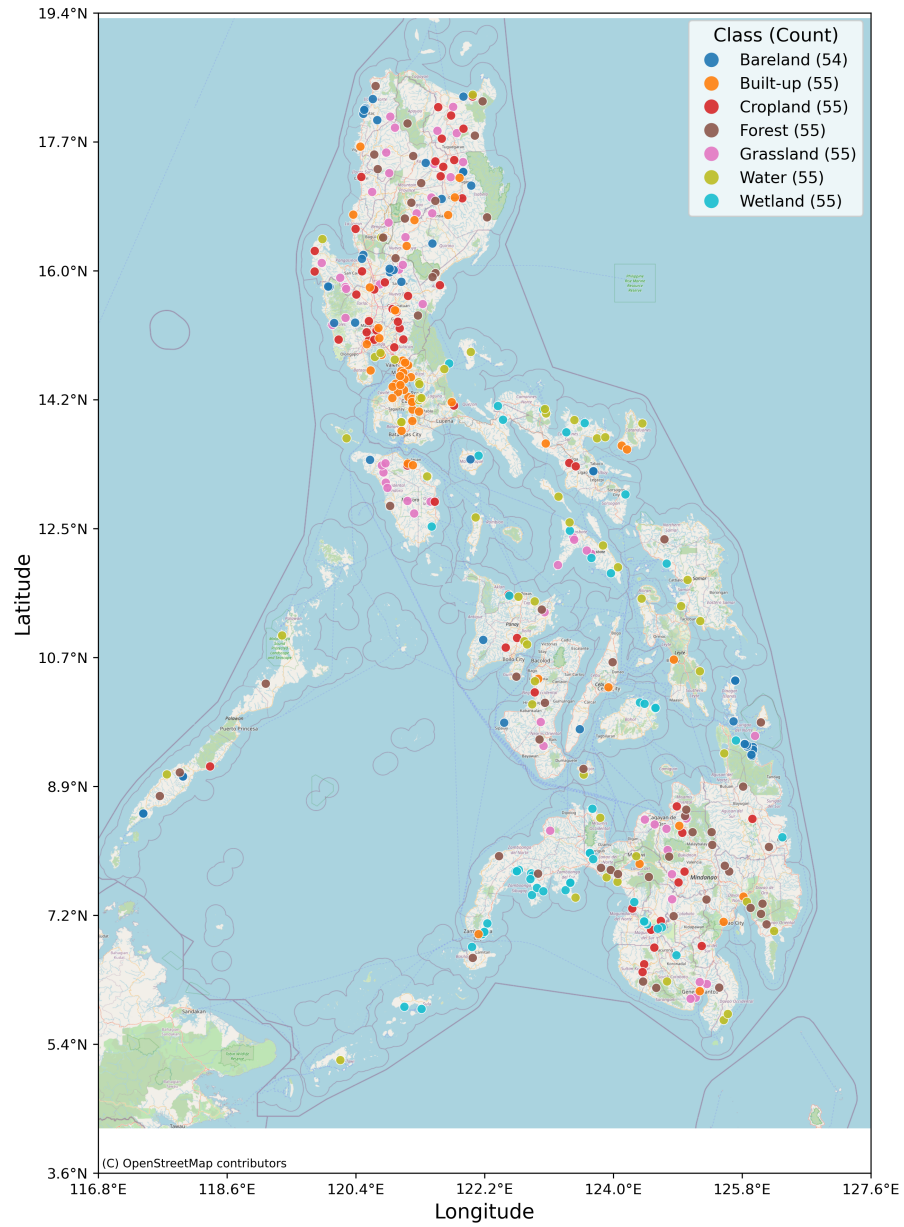


Figure A.8: Spatial Distribution of the Visual Interpretation Samples for Major Agreement ($\tau = 0.7$)

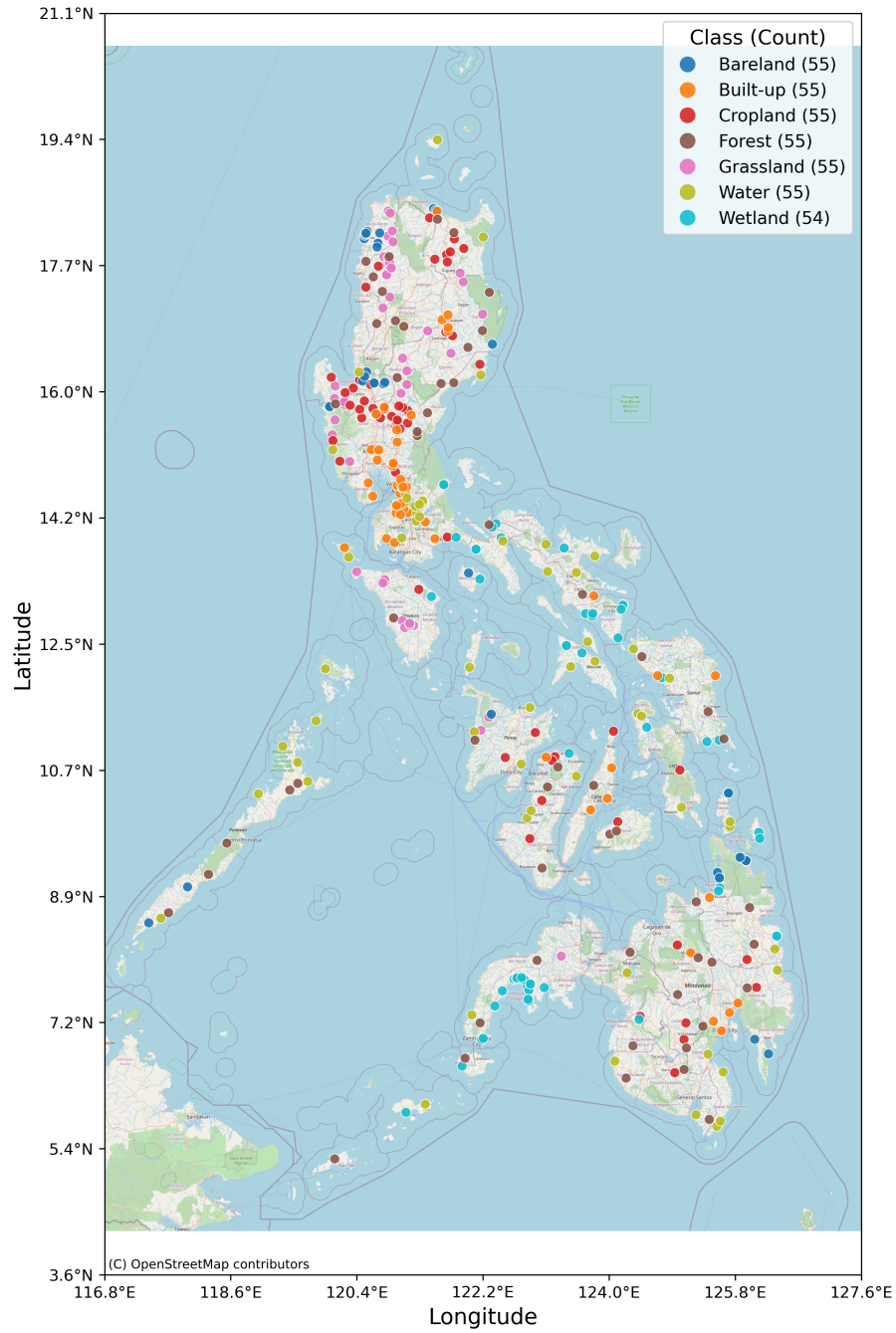


Figure A.9: Spatial Distribution of the Visual Interpretation Samples for Major Agreement ($\tau = 0.8$)

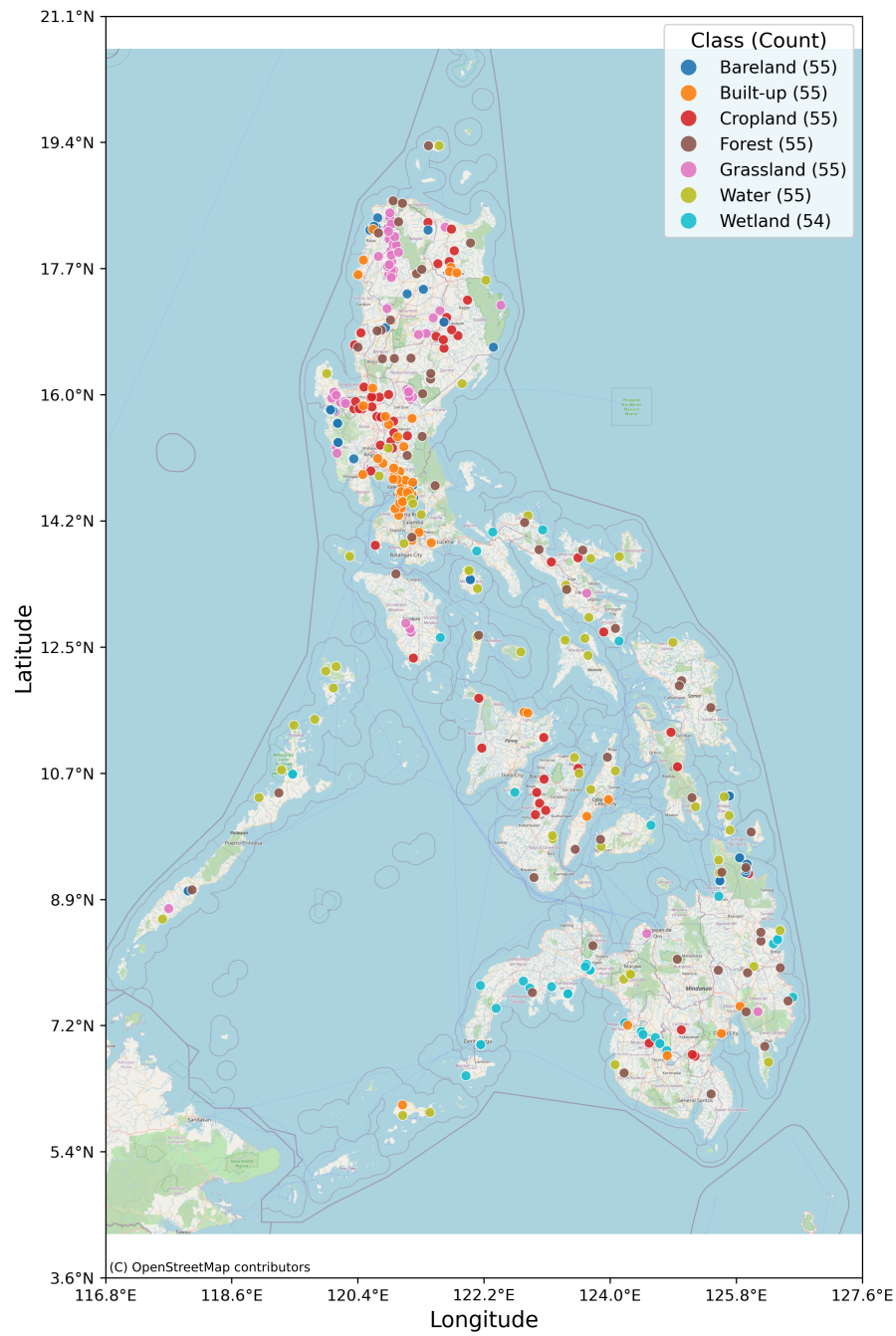


Figure A.10: Spatial Distribution of the Visual Interpretation Samples for Major Agreement ($\tau = 0.9$)

List of Figures

2.1	Study Area	5
2.2	Mapping Matrix of Legend Harmonization	19
2.3	Class Count Distribution of SRS_val	20
2.4	Spatial Distribution of SRS_val	21
2.5	Class Distribution for each Dataset	23
2.6	Flow of Accuracy Assessment	26
3.1	Overall Workflow	28
3.2	Tiling Grid	30
3.3	Overall Workflow of the MOLCA Generation	32
3.4	Temporal Availability of Datasets	34
3.5	Overall Workflow of the MOLCOPA Generation	39
3.6	Overall Workflow of Unsupervised Approach	46
3.7	Auto Labeling Strategy	48
3.8	Overall Workflow of Semi-supervised Approach	49
3.9	Spatial Distribution and Class Count of Set-50	51
3.10	Spatial Distribution and Class Count of Set-100	52
3.11	Spatial Distribution and Class Count of Set-200	53
3.12	MLP Model	57
3.13	Spatial Distribution and Class Count of ValiSet-O	61
4.1	MOLCA 2000	63
4.2	MOLCA 2010	64
4.3	MOLCA 2015	64
4.4	MOLCA 2019	65
4.5	Confusion Matrix for MOLCA 2010 (Interpretation)	68
4.6	Confusion Matrix for MOLCA 2015 (Interpretation)	68
4.7	Confusion Matrices for MOLCA 2019	69
4.8	Overall Agreement	70
4.9	Hybrid Agreement	70

4.10 N-1 Agreement	71
4.11 Major Agreement ($\tau = 0.5$)	71
4.12 Major Agreement ($\tau = 0.6$)	72
4.13 Major Agreement ($\tau = 0.7$)	72
4.14 Major Agreement ($\tau = 0.8$)	73
4.15 Major Agreement ($\tau = 0.9$)	73
4.16 Relationship between Coverage and Overall Accuracy	74
4.17 Overall Accuracy and Kappa Coefficient Comparison (Interpretation) . . .	78
4.18 Accuracy Comparison by Visual Interpretation and SRS_val	78
4.19 Recall per Class by Strategy (Interpretation)	79
4.20 Precision per Class by Strategy (Interpretation)	80
4.21 F1-score per Class by Strategy (Interpretation)	80
4.22 Confusion Matrices for Overall Agreement	81
4.23 Confusion Matrices for Hybrid Agreement	81
4.24 Confusion Matrices for N-1 Agreement	82
4.25 Confusion Matrices for Major Agreement at $\tau = 0.5$	82
4.26 Confusion Matrices for Major Agreement at $\tau = 0.6$	83
4.27 Confusion Matrices for Major Agreement at $\tau = 0.7$	83
4.28 Confusion Matrices for Major Agreement at $\tau = 0.8$	84
4.29 Confusion Matrices for Major Agreement at $\tau = 0.9$	84
4.30 Land Cover Map by KMeans with Global Labeling	85
4.31 Land Cover Map by KMeans with Local Labeling (5-km)	86
4.32 Land Cover Map by KMeans with Local Labeling (10-km)	86
4.33 Confusion Matrices for Local Labeling (10-km)	89
4.34 Confusion Matrices for Local Labeling (5-km)	90
4.35 Land Cover Map by RF with Set-50	93
4.36 Land Cover Map by SVM with Set-50	93
4.37 Land Cover Map by MLP with Set-50	94
4.38 Land Cover Map by MLP with Set-100	94
4.39 Land Cover Map by MLP with Set-200	95
4.40 Confusion Matrices for RF	98
4.41 Confusion Matrices for SVM	99
4.42 Confusion Matrices for MLP	100
4.43 Comparison of Models in Overall Performance	101
4.44 Comparison of Models in F1-Score	102
4.45 Comparison of Models in Recall	103
4.46 Comparison of Models in Precision	104

4.47	Comparison with Existing Datasets in Overall Performance	105
4.48	Comparison with Existing Datasets in F1-Score	106
4.49	Comparison with Existing Datasets in F1-Score (Detailed)	107
A.1	Spatial Distribution of the Visual Interpretation Samples for MOLCA 2010	120
A.2	Spatial Distribution of the Visual Interpretation Samples for MOLCA 2015	121
A.3	Spatial Distribution of the Visual Interpretation Samples for MOLCA 2019 (Overall Agreement)	122
A.4	Spatial Distribution of the Visual Interpretation Samples for Hybrid Agree- ment	123
A.5	Spatial Distribution of the Visual Interpretation Samples for N-1 Agreement	124
A.6	Spatial Distribution of the Visual Interpretation Samples for Major Agree- ment ($\tau = 0.5$)	125
A.7	Spatial Distribution of the Visual Interpretation Samples for Major Agree- ment ($\tau = 0.6$)	126
A.8	Spatial Distribution of the Visual Interpretation Samples for Major Agree- ment ($\tau = 0.7$)	127
A.9	Spatial Distribution of the Visual Interpretation Samples for Major Agree- ment ($\tau = 0.8$)	128
A.10	Spatial Distribution of the Visual Interpretation Samples for Major Agree- ment ($\tau = 0.9$)	129

List of Tables

2.1	Multi-class HRLC Datasets	6
2.2	Binary HRLC Datasets	10
2.3	Accessibility of Datasets	15
2.4	Mapping of MOLCA Classes to FAO LCCS	17
2.5	Legend Harmonization between MOLCA and Existing Datasets	18
2.6	Pixel Count Distribution by Class	24
2.7	Percentage Distribution by Class	24
2.8	Comparison of Overall Accuracy (O.A.) for HRLC Datasets	26
2.9	Comparison of Precision and Recall for Different Land Cover Products	26
3.1	Summary of Target	35
3.2	Summary of Experimental Configurations	55
3.3	Samples' Distribution for MOLCA and MOLCOPA Evaluation	59
4.1	Pixel Count Statistics for each Land Cover Class for MOLCA	66
4.2	Percentage Statistics for each Land Cover Class for MOLCA	67
4.3	Pixel Count Statistics for MOLCOPA-derived LC Maps (Part 1)	75
4.4	Pixel Count Statistics for MOLCOPA-derived LC Maps (Part 2)	76
4.5	Percentage Statistics for MOLCOPA-derived LC Maps (Part 1)	76
4.6	Percentage Statistics for MOLCOPA-derived LC Maps (Part 2)	77
4.7	Pixel Count Statistics for Land Covers Maps by Unsupervised Approach	87
4.8	Percentage Statistics for Land Covers Maps by Unsupervised Approach	88
4.9	Comparison of Overall Accuracy and Kappa Coefficient on ValiSet-O and SRS datasets	92
4.10	Pixel Count Statistics for LC Maps by Semi-supervised Approach	96
4.11	Percentage Statistics for LC Maps by Semi-supervised Approach	96

List of Symbols

Variable	Description
p	Pixel index
c	Land cover class label
i	Index of the input dataset
K	Total number of input datasets
K_{mc}	Total number of multi-class input datasets
K_{bi}	Total number of binary input datasets
$M(i, p)$	Class label of the i -th input map at pixel p
$M_{mc}(i, p)$	Class label of the i -th multi-class map at pixel p
$M_{bi}(i, p)$	Class label of the i -th binary map at pixel p
$N_{mc}(p)$	Count of valid (non-NoData) multi-class maps at pixel p
$\mathbb{I}(\cdot)$	Indicator function
$MOLCA(p)$	MOLCA label at pixel p
$P(c, p)$	Probability of belonging to land cover class c for pixel p
$B(c, p)$	Number of multi-class maps classifying pixel p as class c
$MOLCOPA(c, p)$	MOLCOPA value at pixel p
$P_{max}(p)$	Maximum probability among all classes at pixel p
$C_{cand}(p)$	Candidate class with the highest probability at pixel p
τ	Pre-defined probability threshold
$LC_{OA}(p)$	Land cover label with Overall Agreement at pixel p
$LC_{N-1}(p)$	Land cover label with N-1 Agreement at pixel p
$LC_{major}(p)$	Land cover label with Major Agreement at pixel p
$LC_{hybrid}(p)$	Land cover label with Hybrid Agreement at pixel p

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Vasil Yordanov, for providing support for the completion of this thesis.

My deep and profound thanks go to Prof. Maria Antonia Brovelli. Thank you for your invaluable guidance, continuous support, and immense patience throughout my research. I am truly grateful for the trust you placed in me. I am equally thankful to my co-supervisor, Eng. Qiongjie Xu, for the insightful feedback and expert advice. Your unique perspectives significantly enriched the quality of this work.

To my friends, thank you for being a constant source of joy. The late-night sessions and tough times with you made my journey memorable.

Finally, words cannot express my profound gratitude to my family across the ocean. Thank you for your unconditional support.

Where one journey ends, another begins.

