Enhancing Multi-Resolution Land Characteristics Identification Using Deep Neural Networks for Environmental Monitoring

by Michael Cantrall

B.A. in Business Administration, May 2016, University of Washington
M.S. in Business Analytics, August 2018, University of Washington

A Praxis submitted to

The Faculty of
The School of Engineering and Applied Science
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Engineering

January 9, 2026

Praxis directed by

Jeffrey C. Yu
Professorial Lecturer of Engineering and Applied Science

The School of Engineering and Applied Science of The George Washington University certifies that Michael Cantrall has passed the Final Examination for the degree of Doctor of Engineering as of November 7, 2025. This is the final and approved form of the Praxis.

Enhancing Multi-Resolution Land Characteristics Identification using Deep Neural Networks for Environmental Monitoring.

Michael Cantrall

Praxis Research Committee:

John Fossaceca, Adjunct Professor of Engineering and Applied Science, Praxis Chair

Jeffrey C. Yu, Professorial Lecturer in Engineering and Applied Science, Praxis Advisor

Titus Hei Yeung Fong, Professorial Lecturer in Engineering and Applied Science, Committee Member

© Copyright 2025 by Michael Cantrall
All rights reserved

Dedication

To the author's beloved wife, Lyndsey, whose unwavering love, encouragement, and support have been foundational throughout this journey.

To the author's son, Ethan, whose boundless curiosity and joy serve as a daily reminder of the importance of learning and growth.

To the author's parents, whose sacrifices, guidance, and faith have been the cornerstone of the author's accomplishments.

To the author's sisters, whose love and encouragement have been a constant source of strength and inspiration.

This praxis is dedicated to them with the author's deepest gratitude and love.

Acknowledgements

The author wishes to express heartfelt gratitude to the exceptional professors at George Washington University who guided the cohort and provided the skills and insights essential to reaching this important milestone. Special appreciation is reserved for Dr. Jeffrey C. Yu, whose unwavering support and commitment were invaluable throughout the development of this praxis.

Abstract of Praxis

Enhancing Multi-Resolution Land Characteristics Identification using Deep Neural Networks for Environmental Monitoring

Accurate classification of land use and land cover (LULC) is crucial for effective environmental management, urban planning, and climate modeling. The National Land Cover Database (NLCD) in particular plays a pivotal role as the official source of nationwide land cover data to key federal agencies such as the Environmental Protection Agency (EPA), Federal Emergency Management Agency (FEMA), U.S. Geological Survey (USGS), National Park Service (NPS), U.S. Forest Service, and the National Oceanic and Atmospheric Administration (NOAA). Recent advancements in deep learning, especially transformer-based models, have demonstrated significant potential in improving land use and land cover classifications. These methods have achieved accuracies above 90% on specialized, smaller datasets. While effective, the NLCD classification model is currently achieving classification accuracies around 77.5%.

This praxis harnessed the capabilities of state-of-the-art deep learning models, specifically convolutional neural networks and vision transformers, to improve the accuracy and efficiency of LULC classification across the expansive and multifaceted NLCD dataset. These types of models are well-suited for capturing global dependencies within the data, which is particularly beneficial for the nuanced differentiation required in land LULC classification tasks.

The data was sourced from the Landsat 8 missions at a 30m resolution and paired with 30-500m resolution elevation, ecological, and temporal data. The Landsat imagery

covered nine bands across the visible, near infrared, and short-wave infrared spectra. Enhancements included augmenting the bands to include spectral indices, textural features, temporal mechanisms, and spatial context. The outcome of this research was a significant enhancement in the accuracy of LULC classifications, supporting improved environmental policies and land management practices. By applying newly developed image classification techniques to the precise monitoring of land cover changes, this praxis produced a deep neural network that achieved 92.7% accuracy.

Table of Contents

Dedication iv

Acknowledgements v

Abstract of Praxis vi

Table of Contents viii

List of Figures xiii

List of Tables xiv

List of Equations xvi

List of Symbols xvii

List of Abbreviations and Acronyms xviii

Glossary of Terms 1

Chapter 1—Introduction 2

1.1 Background 2

1.2 Research Motivation 3

1.3 Problem Statement 4

1.4 Thesis Statement 4

1.5 Research Objectives 5

1.6 Research Questions and Hypotheses 6

1.7 Scope of Research 7

1.8 Research Limitations 8

1.9 Organization of Praxis 9

Chapter 2 – Literature Review 10

2.1 Introduction 10
2.2 The Multi Resolution Land Characteristics Consortium and National Land Cover Database 11
2.2.1 Historic Drivers of Land Use Classification 11
2.2.2 Landsat 13
2.2.3 The National Land Cover Database 14
2.3 Spectral Features 15
2.3.1 Raw Spectral Measurements 16
2.3.2 Spectral Independence 17
2.3.3 Spectral Indices 17
2.4 Temporal Consistency and Phenology 18
2.5 Scene Acquisition and Archiving 20
2.5.1 Atmospheric Corrections 20
2.5.2 Scene Retrieval 21
2.5.3 Ancillary Data Sources 22
2.5.4 Sampling and Validation Techniques 23
2.6 Machine Learning Models 23
2.6.1 Tree based Algorithms 25
2.6.2 Convolutional Neural Networks and Deep Learning 25
2.6.3 Single and Multi-Pixel Classifications 27
2.6.4 Semantic Segmented Neural Networks 27
2.6.5 Transformers 29
2.6.6 Unsupervised Classifications 31
2.6.6 Hybrid Neural Network Transformers 31
2.7 Conclusion 32

Chapter 3 – Methodology 35
3.1 Introduction 35
3.1.1 Challenges 36
3.1.2 Methodology Workflow 37
3.2 Data Acquisition and Structure 38
3.2.1 Data Source: Landsat 39
3.2.2 Data Source: NLCD 40
3.2.3 Data Source: DEM 41
3.2.4 Data Source: MODIS 42
3.2.5 Data Source: EPA Ecological Regions 42
3.2.5 Geographic Tiling 43
3.2.6 Temporal Features 43
3.2.7 Data Integration 45
3.3 Data Processing and Exploratory Analysis 46
3.3.1 Adaptive Window Method 47
3.3.2 Data Filtering 49
3.3.3 Grouped Scene Data Split 49
3.4 Feature Engineering and Selection 50
3.4.1 One-Hot Encodings 51
3.4.2 Spectral Indices 52
3.4.3 Textural Features 56
3.4.4 Standardization 57
3.4.5 Feature Selection 58
3.5 Model Training and Optimization 60
3.5.1 Tree-Based Models 62

3.5.2 U-Net 63
3.5.3 Vision Transformer 65
3.5.4 Hybrid CNN + Transformer 66
3.5.5 Hyperparameter Tuning 66
3.6 Model Evaluation 67
3.6.1 Multi-Class Classification Evaluation Metrics 67
Chapter 4 – Results 70
4.1 Introduction 70
4.2 Data Acquisition 71
4.3 Data Preprocessing and Exploratory Analysis 72
4.3.1 Adaptive Window Method 74
4.3.2 Data Filtering 77
4.3.3 Grouped Scene Data Split 78
4.4 Feature Engineering 79
4.4.1 One-Hot Encodings 79
4.4.2 Spectral Indices 81
4.4.3 Textural Features 82
4.4.4 Feature Selection 84
4.5 Model 88
4.5.1 Decision Tree-Based Models 88
4.5.2 U-Net 91
4.5.3 Vision Transformer 94
4.5.4 Hybrid CNN + Transformer 97
4.6 Conclusion 100
Chapter 5 – Discussion and Conclusions 102

5.1 Research Question and Hypothesis 1 102

5.2 Research Question and Hypothesis 2 104

5.3 Research Question and Hypothesis 3 105

5.4 Research Question and Hypothesis 4 106

5.5 Contributions to Body of Knowledge 107

5.6 Recommendations for Future Research 108

References 110

Appendix 124

List of Figures

Figure 3.1.2 – Methodology Overview 37

Figure 3.2 – Raw Data Creation 39

Figure 3.3 – Windowed Sampling Method 47

Figure 3.3.1 - Adaptive Window Method 48

Figure 4.3a - Confidence Bins 73

Figure 4.3b - QA Pixel Category 74

Figure 4.3.1a – Class Confusion 76

Figure 4.3.1b - Class wise Confidence Scores 77

Figure 4.3.3 - Grouped Scene Data Split 78

Figure 4.4.1 - EPA Ecoregions 80

Figure 4.4.2 - Normalized Vegetation Index (NDVI) 82

Figure 4.4.3 - Gray-Level Co-Occurrence Sample Values 83

Figure 4.4.4a - Principal Component Analysis 84

Figure 4.4.4c - Gini Importance 86

Figure 4.5.1 - Random Forest Confusion Matrix 91

Figure 4.5.2a - U-Net Forest Class Confusion Matrix 93

Figure 4.5.2b - U-Net Loss and Accuracy 93

Figure 4.5.3a - SegFormer Confusion Matrix 96

Figure 4.5.3b - SegFormer Loss and Accuracy 97

Figure 4.5.4a - STransU2Net Confusion Matrix 99

Figure 4.5.4b - STransU2Net Loss and Accuracy 100

Figure 5.1 – Intra-Class Confusion for Forest Class 103

List of Tables

Table 2.6 - NLCD Model History 24

Table 2.7 - Literature Review Key Findings 33

Table 3.1.1 - NLCD Classifier Challenges 36

Table 3.2.1 – Landsat Bands 40

Table 3.2.2 – NLCD Data 41

Table 3.2.3 – DEM Data 41

Table 3.2.4 – MODIS Data 42

Table 3.2.5 – EPA Ecological Regions 43

Table 3.2.3 - MODIS Phenological Dates Sample 44

Table 3.4.2a – Symbols and Definitions for Index Calculations 53

Table 3.4.2b – Tasseled Cap Weights 53

Table 4.2a - Data Acquisition 71

Table 4.2b - Land Use Classification 72

Table 4.3.1 – Adaptive Window Results 75

Table 4.4.1 – Quality Assurance One Hot Encoded Layers 80

Table 4.4.2 – Spectral Index Features 81

Table 4.4.4a – Binary Class SHAP Importance 85

Table 4.4.4b – Binary Class Gini Importance 86

Table 4.4.4 U-Net vs SegFormer Feature Importance 87

Table 4.5.1a Random Forest Hyperparameters 89

Table 4.5.1b XGBoost Hyperparameters 89

Table 4.5.1c Tree-Based Model Outputs 90

Table 4.5.1d Random Forest Temporal Results 90

Table 4.5.2a U-Net Hyperparameter Search 92

Table 4.5.2b U-Net Temporal Results 92

Table 4.5.3a SegFormer Hyperparameter Search 94

Table 4.5.3b SegFormer Temporal Results 95

Table 4.5.4 STransU2Net Hyperparameter Search 98

Table 4.5.4b STransU2Net Temporal Results 98

Table 4.6 – Model Experiment Comparison 101

Table 5.3 - STransU2Net vs U-Net Performance 105

Table 5.4 - Temporal Accuracy Comparison Across Models 106

Table A1 – QA_PIXEL Bit Values 124

Table A2 – QA_RADSAT Bit Values 125

Table A3 – QA_AEROSOL Bit Values 126

Table A4 – NLCD Land Use Classes 127

Table A5 – 2019 NLCD Accuracy Assessment 129

List of Equations

Equation 1 – Tasseled Cap Calculation .....**Error! Bookmark not defined.**

Equation 2 – Normalized Difference Vegetation Index ....**Error! Bookmark not defined.**

Equation 3 – Enhanced Vegetation Index..... 53

Equation 4 – Chlorophyll Index Green..... 54

Equation 5 – Normalized Difference Water Index 54

Equation 6 – Normalized Difference Snow Index..... 54

Equation 7 – Normalized Difference Building Index..... 55

Equation 8 – Modified Soil-Adjusted Vegetation Index 55

Equation 9 – Normalized Difference Moisture Index..... 55

Equation 10 – Gray Level Co-Occurrence Matrix Contrast..... 56

Equation 11 – Gray Level Co-Occurrence Matrix Dissimilarity..... 56

Equation 12 – Gray Level Co-Occurrence Matrix Homogeneity..... 58

Equation 13 – Gray Level Co-Occurrence Matrix Energy 56

Equation 14 – Gray Level Co-Occurrence Matrix Correlation 57

Equation 15 – Shapley Additive Explanation..... 58

Equation 16 – Sample Weights..... 62

Equation 17 – Overall Accuracy..... 67

Equation 18 – Accuracy..... 68

Equation 19 – Balanced Accuracy..... 68

Equation 20 – Mean Intersection Over Union 68

List of Symbols

mIoU	Mean Intersection over Union
OA	Overall Accuracy
ρ_B	Blue reflectance
ρ_G	Green reflectance
ρ_R	Red reflectance
ρ_{NIR}	Near-Infrared reflectance
ρ_{SWIR1}	Shortwave Infrared 1 reflectance
ρ_{SIWR2}	Shortwave Infrared 2 reflectance
ρ_{TIR}	Thermal Infrared 1 (surface temperature)
y	Ground truth labels
$\hat{y}$	Predicted Label

List of Abbreviations and Acronyms

ARD	Analysis Ready Data
BAEI	Built-up Area Extraction Index
BLM	Bureau of Land Management
BRBA	Band Ratio for Built-up Area
CI_green	Chlorophyll Index (Green)
CONUS	Conterminous United States
EPA	Environmental Protection Agency
EVI	Enhanced Vegetation Index
GIS	Geographic Information System
LULC	Land Use Land Cover
MNDWI	Modified NDWI
MRLC	Multi-Resolution Land Characteristics
MSAVI	Modified Soil-Adjusted Vegetation Index
NBI	New Built-up Index
NDBI	Normalized Difference Built-up Index
NDMI	Normalized Difference Moisture Index
NDSI	Normalized Difference Snow Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NLCD	National Land Cover Data
NOAA	National Oceanic and Atmospheric Administration

NPS	National Park Service
TCB	Tasseled Cap Brightness
TCG	Tasseled Cap Greenness
TCW	Tasseled Cap Wetness
UI	Urban Index
USDA	United States Department of Agriculture
USFS	United States Forest Service
USFWS	U.S. Fish and Wildlife Service
USGS	United States Geological Survey

Glossary of Terms

Analysis Ready Data (ARD)	Satellite imagery that has been preprocessed to correct for distortions, atmospheric effects, and other sensor noise, which makes it ready for analysis without requiring additional manipulation.
Area of Interest (AOI)	A predefined geographic boundary used to extract and process remote sensing data.
Band	A single channel of satellite sensor data that records reflectance within a specific range of the electromagnetic spectrum. Multiple bands are used together to analyze surface characteristics.
Ecoregion	A geographically distinct area defined by shared environmental conditions such as climate, vegetation, geology, or human interactions.
Land Cover	The physical material present on Earth's surface, such as forests, water, wetlands, or urban areas.
Land Use	The human interaction in a given area of land they occupy, such as urban development, cultivated crops, and pasture land.
Landsat	A series of Earth observation satellites operated by the U.S. Geological Survey and the National Aeronautics and Space Administration.
Layer	A single dataset representing one variable, such as surface reflectance, temperature, or elevation.
Phenology	Seasonal cycles in vegetation, such as spring “green-up” and fall “dormancy”.
Reflectance	The proportion of the electromagnetic spectrum that reflects off the Earth’s surface and is captured by remote sensors.
Remote Sensing	The observation of Earth’s surface from a distance, often via satellites or aircraft.
Scene	A single satellite image capture of a specific geographic area, which could include multiple bands or layers.

Chapter 1—Introduction

1.1 Background

In today's interconnected world, comprehensive and precise information across various interrelated domains is essential for effective decision-making. Land use itself stands at the core of addressing a vast array of critical challenges, such as unplanned urban expansion, environmental degradation, the depletion of agricultural lands, the destruction of wetlands, and the erosion of biodiversity. The ability to understand and manage land use dynamics is and will be integral in the mitigation of climate impacts, natural resource allocation, ecosystem preservation, and sustainable growth and development (Anderson et al., 1976). In 1966, the Secretary of the Interior Stewart Udall announced the project Earth Resources Observation Satellites (EROS) by the United States Geological Survey (USGS) in collaboration with the National Aeronautics and Space Administration (NASA) aimed at using earth-orbiting satellites carrying remote sensing instruments to gather facts about the natural resources of earth (Udall, 1966).

The EROS project was the predecessor of the 1972 Earth Resource Technology Satellite (ERTS)-A Project. ERTS-A was the first satellite launched specifically for the remote sensing of Earth's resources. It reduced the imagery requirements for a full continental United States view from 500,000 high altitude aircraft images to 500 from ERTS-A providing global coverage every 18 days (McRoberts & Lynch, 1972). ERTS-A was later renamed Landsat 1 and started the collection of what is now referred to as Landsat remote sensing data.

In 1984, the Land Remote Sensing Commercialization Act initiated the commercialization of Landsat remote sensing data, resulting in significant cost increases and restrictions on data distribution. By the late 1980s, the price of acquiring a single moment in time snapshot for each of the 430 path/rows covering the continental United States (CONUS) would cost $1.72 million USD (Wickham et al., 2014). Out of this need to reduce costs, the Multi-Resolution Land Characteristics Consortium (MRLC) was created in the mid-1990s to compile comprehensive national Landsat data by the USGS, NOAA, and the EPA. The MRLC collaboration led to the development of the National Land Cover Database (NLCD), which has become established as the *de facto* standard dataset for environmental monitoring, planning, and management across various sectors due to its accuracy, standardization, and affordability.

1.2 Research Motivation

Periodic assessments of land use and land cover (LULC) classification models are crucial for ensuring the accuracy of many downstream activities. These assessments help in identifying methodological strengths and architectural changes that pinpoint areas needing improvement, guiding future enhancements and resource allocation. The accuracy of LULC classification models has remained a persistent challenge for the MRLC despite extensive advancements in methodologies. Early assessments, such as the thematic accuracy evaluation of the 1992 NLCD, revealed significant limitations, including an overall accuracy ranging from 43% to 83% depending on the class (Stehman et al., 2003). Modern level accuracy was achieved with the 2001 NLCD with the adoption of decision tree models, as accuracy climbed to 78.7% for level 1 classification (J. D. Wickham et al., 2010). In 2019, random forest models were adopted, and the

models increased to 83.1% for level 1 and 77.5% for level 2. In 2024, the NLCD moved toward leveraging U-Net models for improved classification, although no published accuracy assessment has been conducted. (Brown, 2024). The timing of accuracy assessments remains a challenge, often averaging 3-4 years after the data was created, as highlighted by the 2023 accuracy assessment conducted on the 2019 NLCD.

1.3 Problem Statement

NLCD classification has not kept pace with escalating demand for highly accurate, authoritative land use and land cover data (Wickham et al., 2023)*, undermining the NLCD's reliability for land management and climate modeling* (Arpitha et al., 2023) .

The MRLC has made extraordinary achievements in the development of the NLCD through additional feature sets, ancillary data, and overall standardization of the process. Despite these remarkable strides, achieving highly accurate and consistent models remains a complex challenge. The dimensionality of the data itself is so incredibly vast, accounting for spatial, temporal, and spectral variability. This includes a number of factors such as elevation, soil types, vegetation types, cloud cover, sun angles, along with countless other features that may yet be uncovered. These models play a critical role in addressing some of the most pressing challenges facing society today, from climate change mitigation to sustainable urban planning and the conservation of our vulnerable ecological systems.

1.4 Thesis Statement

Deep neural network models with attention mechanisms are required to improve NLCD land use classification.

Deep neural networks excel at handling the inherent complexity of land use classification by effectively modeling the high-dimensional and heterogeneous nature of the data, as seen by the recent adoption of U-Net architecture by the MRLC. More specifically, attention mechanisms significantly enhance the DNN's capabilities by capturing both local and global dependencies in the data. This allows for the model to dynamically focus on the most relevant features while maintaining a holistic understanding of the broader context as an essential requirement for high accuracy in distinguishing complex and overlapping land features.

1.5 Research Objectives

This praxis aimed to contribute to the ongoing efforts of the MRLC toward providing a comprehensive, singular dataset that can be leveraged for land use by advancing machine learning methodologies. Enhancing classification accuracy can enable more reliable insights for urban planning, natural resource management, and climate impact assessments.

The research is guided by the following specific objectives:

- Examine the limitations of current land use classification methods in accurately capturing land features.
- Investigate the role of attention mechanisms in enhancing model performance for land cover types.
- Develop a predictive model that leverages advanced deep learning architectures to achieve a 90% accuracy in land use classification, a 12.5% improvement.

Achieving these objectives would make a significant contribution to the field of geospatial data science, providing a framework for predictive modeling in land use applications.

1.6 Research Questions and Hypotheses

As this research is focused on developing a deep neural network model for improving LULC classification in CONUS, it is critical to analyze the comparative performance of advanced architectures and mechanisms to enhance classification accuracy. The latest thematic assessment of the NLCD was completed using a random forest model, which served as the baseline for this research. While no thematic assessment has been conducted, the latest user guides from the MLRC indicate they are currently using U-Net models for classification. While no published accuracy rating exists, U-Nets are essential in validating the most accurate models. Specifically, this praxis explores the role of random forest, U-Net, Vision Transformers (ViT), and temporal mechanisms.

This praxis seeks to address the following research questions, each accompanied by a descriptive short name used throughout the text:

- **RQ1**: *Deep networks for forest accuracy gains* - Do deep neural networks outperform random forest models in differentiating forest types in the continental United States?
- **RQ2**: *Vision transformers for accuracy gains* - Do vision transformers outperform U-Net in LULC classification?

- **RQ3**: *Hybrid Transformers for accuracy gains* - Do hybrid CNN + transformer models outperform U-Net in LULC classification?
- **RQ4**: *Temporal context for accuracy gains* - Does incorporating temporal mechanisms reduce classification errors in land cover mapping in level 2 areas?

Based on the research questions outlined above, the following hypotheses are formulated:

- **H1**: *Deep networks for forest accuracy gains* - Deep neural networks will outperform random forest models in accurately detecting forest types by 10% in the continental United States.
- **H2**: *Vision transformers for accuracy gains* - Vision transformers will outperform U-Net in LULC classification by 5% in accuracy.
- **H3**: *Hybrid Transformers for accuracy gains* - Hybrid CNN + transformers will outperform U-Net in LULC classification by 5% in accuracy.
- **H4**: *Temporal Context for accuracy gains* - Time-series analysis enhances the detection accuracy of LULC in both level 1 and level 2 classification by 2%.

For clarity in the remainder of this praxis RQ1-RQ3 and H1-H3 are collectively referred to as the model architecture comparisons and represent the model-based objectives.

1.7 Scope of Research

This praxis focuses on developing a deep neural network model capable of detecting multi-resolution land characteristics across the continental United States. The study emphasizes differentiating Level 1 categories, such as forest, agricultural land, water, and urban development, as well as Level 2 subcategories to capture land use and land cover from 2019. The research leverages USGS analysis-ready data (ARD) and the

NLCD, integrating these rich sources of spatial, temporal, and multi-band spectral data. This research focuses on designing and training deep learning models integrated with attention mechanisms. The models undergo rigorous testing to identify the most effective architecture, with the objective of achieving a 12.5% improvement over current classification accuracies.

1.8 Research Limitations

The MRLC employs a rigorous set of processes to prepare, validate, and refine its datasets, including the preprocessing of raw satellite imagery, data validation, and post-classification cleanup to ensure the quality and reliability of NLCD datasets. These processes, which include steps such as Continuous Change Detection (CCD), cloud filtering, and shadow filtering, are labor-intensive and require significant time and resources. This research, however, relies solely on ARD provided by Landsat. As a result, the study does not involve the processing or post-processing stages, which can take months of work by specialized teams. The focus remains strictly on the evaluation of the proposed deep learning models for land classification.

Exact copies of the remote sensing data used for the NLCD products for 2019 through 2024 are not downloadable. While the source of the data is the same, the exact frame from the exact moment in time used to produce the classification may vary. The MRLC uses all remote sensing data, then preprocesses it, while the ARD may use data from a different point in time, causing discrepancies between the training data and labeled datasets. This effect is mitigated in the methodology, but remains an issue. This praxis aligns closely with the MRLC's existing work by adhering to established land

classification categories and CONUS maps, ensuring compatibility and enhancing its potential for future integration within their framework.

1.9 Organization of Praxis

The remainder of this praxis is organized in the following manner:

Chapter 2, The Literature Review, provides a comprehensive review of existing research, focusing on the application of the NLCD dataset and MRLC consortium products for land cover classification and change detection, while highlighting gaps in leveraging advanced machine learning for environmental monitoring. Chapter 3 details the datasets used, including NLCD products and Landsat 9 remote sensing imagery, and outlines the development and comparison of deep neural network models for improving land cover classification and identification. Chapter 4 presents findings from the application of the models, including assessments of their performance and accuracy in enhancing multi-resolution land characteristics identification. Chapter 5, the discussion and conclusions, summarizes key findings and their significance, discusses contributions to the field, addresses the research questions, and explores potential avenues for future research.

Chapter 2 – Literature Review

2.1 Introduction

This praxis advances the research of land use and land cover (LULC) classification by applying state-of-the-art computer vision and machine learning methods to lift the National Land Cover Database (NLCD) past its long ceiling of 77.5% accuracy. LULC classification is a foundational element of environmental management, urban planning, agriculture, disaster mitigation, and land policy. Despite significant advancements in development, high classification accuracy in large-scale LULC applications such as the U.S. NLCD remains a challenge with the 2023 accuracy assessment of the 2019 NLCD at 77.5% overall accuracy, only 5-6 percentage points better than the 2001-2016 releases. Current gaps in the accuracy of remote sensing imagery classification present opportunities to leverage state-of-the-art advancements in spectral data analysis, temporal choices that govern scene selection, integration of ancillary layers that provide contextual information, and the machine learning models - from the tree-based algorithms used throughout NLCD's past to convolutional neural networks (CNNS) and vision transformers (ViTs). The following literature review is organized into the following sections:

- §2.2: The Multi-Resolution Land Characteristics Consortium and the National Land Cover Database Production workflow
- §2.3: Spectral features: Raw image bands, indices, and pixel-based challenges
- §2.4: Temporal consistency and phenology
- §2.5: Scene acquisition: atmospheric corrections, ancillary data, archiving, and sampling techniques

- §2.6 Machine learning models: historic models to state of the art

The conclusion section (§2.7) summarizes the open research gaps and research opportunities from the key research literature that justify the methodology of this praxis.

2.2 The Multi-Resolution Land Characteristics Consortium and National Land Cover Database

The National Land Cover Database (NLCD) serves as the definitive land cover database for the United States, supporting federal, state, local, and nongovernmental applications that seek to gain insights into ecosystems, biodiversity, climate change effects, and land management (C. A. Barnes, 2014). The NLCD is produced by the Multi-Resolution Land Characteristics (MRLC) Consortium. The MRLC was founded in 1992 by the U.S. Geological Survey (USGS), the National Oceanic and Atmospheric Administration (NOAA), and the U.S. Environmental Protection Agency (EPA) and has since then extended to include the U.S. Department of Interior (DOI), Department of Commerce (DOC), U.S. Department of Agriculture (USDA), U.S. Forest Service (USFS), Bureau of Land Management (BLM), National Park Service (NPS), Fish and Wildlife Service (USFWS), Department of Defense (DOD), and U.S. Army Corps of Engineers (USCOE) (Wickham et al., 2014). The creation of the NLCD is led by a science team within the USGS under the Land Cover program using the Landsat Analysis Ready Data (Brown, 2024) captured from Landsat satellites by the USGS Earth Resources Observation and Science Center (EROS).

2.2.1 Historic Drivers of Land Use Classification

To understand the breadth and depth of the output of LULC classifications, its evolution must be traced over time back to humanity's quest for a true bird's-eye view.

Absence of empirical data in early cartography led to increasingly important errors that were often reproduced simply because replication of inaccuracies were easier than excising them (Fernández-Armesto, 2011). Starting in the 1500s, empirical surveys helped remove some of those inaccuracies, but still often misregistered coastlines and mismatched class labels (Alexander, 2023). As pilots took to the sky in tethered-balloon photos, during Wright-brother flyovers, and while executing World War reconnaissance missions, draftsmen shifted to real imagery via aerial photography from surveyor measurements, resulting in more accurate measurements (J. Campbell & Wynne, 2011). However, every flight changed camera lenses, film, altitude, and time of year, producing spectral and temporal inconsistencies needed for a large-scale application (J. B. Campbell et al., 2023).

The first national scale land use inventory for the U.S. was completed by Francis J. Marschner in the 1940s by assembling 1:1,000,000 scale mosaics of aerial photographs taken throughout a decade during the 1930s-1940s revealing how labor-intensive this task truly was (Anderson et al., 1976). Post-World War II era aerial surveillance saw the adoption of infrared film, increased camera quality, and flight frequencies, leading to increased accuracy and spectral analysis. However important this work was, it highlighted the labor-intensive constraints that would remain until the 1960s, when the U.S. National Aeronautics and Space Administration (NASA) established a research program dedicated to remote sensing, and the U.S. National Academy of Sciences (NAS) specifically researched remote sensing for agriculture and forestry (J. Campbell & Wynne, 2011).

In 1971 the formation of an Inter-agency Steering Committee on Land Use Information and Classification was formed to meet the needs of Federal agencies to maintain a broad overview of land use patterns, trends, and environmental impacts. The committee was composed of 153 representatives from the Geological Survey of the U.S. Department of the Interior, the Earth Observations Program of NASA, the Soil Conservation Service of the U.S. Department of Agriculture, the Association of American Geographers, the International Geographical Union, and countless other regional, State, local government agencies, universities, institutes, and select private partners (Gerlach et al., 1971). The committee was the first of its kind for land use and land change due to the diversity of attendees, all with the common goal of developing a national classification system that would standardize data (Anderson et al., 1976). While many Federal, State, and local agencies were collecting data about land, they were often working independently, resulting in a duplication of effort and differences in definitions and categorizations to the point that it was deemed nearly impossible to reconcile. The most important outputs of the collection of representatives are the agreements of level 1 and level 2 land classifications. These classifications—such as water, snow, forest, urban development, wetland, cultivated land, herbaceous, barren, and shrubland—remain the backbone of the NLCD (Anderson et al., 1976).

2.2.2 Landsat

In 1966 project EROS (Earth Resources Observation Satellites) by the USGS in collaboration with NASA aimed at using earth-orbiting satellites carrying remote sensing instruments (Udall, 1966). In 1972, the launch of Earth Resource Technology Satellite 1 (ERTS-1) reduced the number of high-altitude aircraft images to cover the continental

United States (CONUS) from about 500,000 to just 500 satellite images. This new satellite could provide global coverage every 18 days (McRoberts & Lynch, 1972). Over the decades, the Landsat program expanded from Landsat 1 to Landsat 9, launched in 2021, with increasingly accurate sensors and more frequent data collection (Sayler, 2022). Today, Landsat imagery is the standard used by the MRLC to produce the NLCD (Brown, 2024).

2.2.3 The National Land Cover Database

Since its creation in 1992, MRLC has produced annual land cover data for the years 1985-2023. There have been six NLCD accuracy assessments conducted for the 1992, 2001, 2006, 2011, 2016, and 2019 NLCDs. These assessments show a consistent usage of tree-based machine learning models and a range of methodologies, achieving accuracy scores approaching 80%. In 2021, the MRLC began using an ensemble U-Net model, but no official accuracy assessment has yet been conducted (Brown, 2024). The MRLC has made great strides in standardizing land use and land cover classification, automation, and assessments. However, new technologies that offer a chance of producing the NLCD to a higher degree of accuracy, better capturing local environmental features as well as global dependencies.

The NLCD itself has grown into an impressively used source of information, being cited in 9,165 journal articles in 1,085 different journal sources, 516 conference papers, 111 U.S. policy sources, over 10,000 news mentions in 69 different countries, 1300 state, federal, and local technical reports, and used in over 900 theses and dissertations (EROS Center, 2024). The NLCD is used to derive the annual U.S. greenhouse gas emission estimates reported to the United Nations Framework

Convention on Climate Change (UNFCCC) (Wickham et al., 2023). AECOM is a private company that has utilized the NLCD for environmental modeling for EPA violations (Warren, 2022). Tetra Tech, a consulting firm, worked on pollution modeling for the USGS and the EPA utilizing the NLCD to differentiate types of pollution based on the land use (Sievers, 2014). The National Academies of Science, Engineering, and Medicine have used the NLCD as part of their research on the water quality in Coeur d'Alene, Idaho by using historic land use classifications to identify sources of high metals contamination (Committee on the Future of Water Quality in Coeur d'Alene Lake et al., 2022). To ensure that policy, land management, and environmental services are accomplishing their individual missions, it is imperative that the NLCD strives to achieve the highest accuracy it can. Even a small increase in accuracy could be felt beyond the remote-sensing world and into cleaner regulations, better hydrology models, and an overall better understanding of how our world is changing.

2.3 Spectral Features

Spectral features consist of spectral bands that are raw recordings from sensors and their derivatives (Karakizi et al., 2018), this information makes up the majority of the input data for land use and land cover classification algorithm. Regarding LULC classification, specifically, the transformation of spectral input features can lead to increased spatial and spectral value meaning (Lee et al., 2020). Each input set, whether a single pixel or a set of pixels is made up of layers of spectral values and is responsible for the distinguishing of the different NLCD classes. Overall classification accuracy is directly tied to these spectral features, an understanding of these basic spectral features and a comparison of the derivatives is necessary to create a well performing model. For

the latest NLCD thematic assessment the overall accuracy of identifying certain forest, barren, and wetlands have a user accuracy of lower than 60% (Wickham et al., 2023). In the 1992 the MRLC realized that most classes fall on a spectral continuum, a forest area being marked as deciduous or evergreen when it has a mixed forest, or low and medium residential density would be hard to accurately define edge cases (Vogelmann, 2001).

2.3.1 Raw Spectral Measurements

The Landsat images produced are referred to as surface reflectance, which is a recording of solar radiation reflecting off of Earth's surface and primarily consists of visible and near-infrared spectrums (J. Campbell & Wynne, 2011). Objects emit electromagnetic radiation and reflect radiation that has been emitted by other objects. In the case of satellite imagery, two types of reflectance exist: Top-of-atmosphere (TOA) and Bottom-of-atmosphere (BOA) reflectance. TOA corrects for sensor or band sensitivity and solar illumination but does not make atmospheric corrections. BOA is also known as surface reflectance (SR) and removes atmospheric effects (J. B. Campbell et al., 2023). One of the largest adjustments to the 2019 NLCD was the use of surface reflectance data rather than top-of-atmosphere reflectance, which helps with atmospheric effects (Wickham et al., 2023).

The most common form of electromagnetic radiation is visible light, which is a limited portion of the electromagnetic spectrum. Visible light is captured using red, green, and blue (RGB) sensors. Just outside the visible spectrum exists a near-infrared region (NIR), which is primarily used for capturing vegetation. The visible and near infrared spectrum (VNIR) contains both visible light and NIR. The shortwave infrared region (SWIR) is not visible but can capture minerals, fires, crop health, and surface

moisture. Thermal radiation captures temperature information. These encapsulate everything within Landsat scenes; however, near visible light is also gamma rays, x-rays, ultraviolet light, microwave, and radio radiation (J. B. Campbell, 2011).

2.3.2 Spectral Independence

When designing an algorithm for LULC, it is important to keep a many to one relationship between spectral classes and the information class. Spectral classes are groups of pixels that align in numeric recordings directly from remote sensed image bands. It is expected to have variation in our spectral classes with respect to the information class, i.e., many species, density, height, and foliage in a forest class. If the spectral class belongs to many informational classes this will decrease the accuracy of the algorithm (J. B. Campbell et al., 2023).

2.3.3 Spectral Indices

One of the oldest and most common forms of feature engineering for LULC is spectral indices, which create highly consistent information expansions highlighting specific classes (X. Zhao et al., 2022). Spectral indices are the calculation of two or more image bands linked to create a spectral features (Sabr, 2024). The following list shows the most common indices used today:

- *Normalized Difference Vegetation Index* (NDVI) has been used to calculate vegetation status, growth, and the removal of vegetation (H. Zhao et al., 2018). NDVI was utilized in the creation of the 2001 NLCD (Limin Yang et al., 2001).
- Normalized Difference Water Index (NDWI) has been used to identify precipitation, groundwater volume, and waterbodies (Ling et al., 2022).

- Normalized Difference Building Index (NDBI) has been used as an estimator of building existence (Ling et al., 2022).

2.4 Temporal Consistency and Phenology

Recent research shows that the value of time series observations has helped with understanding construction, agricultural activities, forest management, flooding, fire, and the regrowth of forest lands (Fu et al., 2024). Frequent adjustments are made to the NLCD via a leaf-on, leaf-off strategy, which helps highlight vegetation changes. Temporal analysis is used today in the collection 1 of the NLCD to detect spectral changes from year to year, where a lack in spectral change indicates congruency within the classification of a pixel window (Brown, 2024, p. 1). These temporal features are attributed to a 1-2% increase in overall accuracy (Jin et al., 2023). Leaf-on leaf-off strategy creates a temporal dataset that attempts to capture peak and minimal vegetation normally in spring/summer and fall/winter months. One of the problems this strategy helps with is the difficulty in recognizing deciduous versus coniferous vegetation (Vogelmann, 2001). It also improves the recognition of some agriculture lands, since crops are often planted and harvested in the same year, and then lie barren during winter months. Some leaf-on, leaf-off strategies include tight temporal windows which subsequently helped with cloud, shadow, and other seasonal variabilities that existed with year to year analysis (Jin et al., 2023).

The first NLCD in 1992 used a leaf-on or leaf-off strategy to capture deciduous plants well (Vogelmann, 2001). In 2001 instead of relying on a leaf-on or leaf-off strategy, it consisted of three dates for early, peak, and late growing seasons (Limin Yang et al., 2001). There was a total of 16 mapping zones, all of which had different scene

dates; for example, zone 1 covered from Southern Idaho to Northern Nevada. The scene period for zone 1 was June 19th to July 16th for maximum NDVI leaf on period, Sept 12th to October 9th for low NDVI leaf off period, followed by April 9th to May 7th for a second low NDVI leaf off period (Limin Yang et al., 2001). In 2006 the NLCD images were taken within two weeks of the anniversary of the 2001 images, and only using a single leaf-on scene for each target. However, they implemented a year over year comparison using the 2001 pixels as a baseline, if pixels from 2001 to 2006 retained their spectral imaging the label classes from 2001 were also retained (J. D. Wickham et al., 2013).

New advancements in temporal transformations of spectral imagery for land cover classification offer the possibility for additional accuracy in low accuracy classes for the NLCD. Temporal elements historically are submitted as different layers translated into a single dimensional vector, while optionally creating line and polygon representations have seen increased accuracy in convolutional neural networks with low samples (Lee et al., 2020). The line plot maps temporally on the x-axis and the spectral bands on the y-axis, showing temporal spectral changes throughout the year. The polygon representation is a radial chart showing the seasonal changes as a counterclockwise quadrant with the spectral bands radiating from the center also showing the temporal spectral changes throughout the year.

The temporal phenological data records from NASA are the Moderate Resolution Imaging Spectroradiometer (MODIS) MCD12Q2 dataset. MODIS records the timing of phenometric vegetation cycles using an Enhanced Vegetation Index (EVI2) from first appearance of green growth through the growing season and into dormancy as Greenup, MidGreenup, Maturity, Peak, Senescense, MidGreendown, and Dormancy (Gray et al.,

2019). This data is offered in 500 meter grid from 2001 until at least 2025. This data can capture rapid land changes during the growing season and offers a high temporal resolution at 1-2 day intervals (Gu et al., 2016). The MODIS MCD12Q2 offers higher temporal and finer geospatial granularity than the current NLCD phenological window methods.

2.5 Scene Acquisition and Archiving

Outside of spectral and temporal consistency getting an image without distortions is a difficult task. For the 2016 NLCD only 70% of selected tiles were free of clouds, 25% were free of clouds within 1 year of 2016, and 5% had to be outside the year window (Jin et al., 2023). Most fixes include noise removal, gap-filling techniques, cloud positioning and size (Jin et al., 2023). The 1992 NLCD utilized Digital Terrain Elevation Data, census population density, and National Wetlands Inventory, as well as derived data from State Soil Geographic Base, available water capacity, and various state or national programs (Stehman et al., 2003).

2.5.1 Atmospheric Corrections

When imagery is acquired with a sensor from Landsat it is recorded as a scene (J. B. Campbell, 2011). Every scene is calibrated using radiometric and geometric correction and terrain-orthorectification, so pixels are free from common distortions and are consistent. Level 1 data goes through a quality assessment pixel level band (QA_PIXEL) which flags clouds, shadow, ice, and water along with their confidence levels to filter and allow for masking. Level 1 also includes a Quality Assessment Radiometric Saturation and Terrain Occlusion Band (QA_RADSAT), which tells us when sensors are saturated or not usable. The Level 2 data includes processing that converts Top-of-Atmosphere

values to surface reflectance (LaSRC) and surface temperature which removes atmospheric noise from surface level reflectance. Lastly, both the primary and ancillary provided data are stored into Mission Data streams and scheduled for collection, these are handled using lossless compression (Sayler, 2022). NLCD use of Landsat Collection Level 2 Analysis Ready Data (ARD), which provides good retrieval methods (Brown, 2024, p. 1). The Collect 2 ARD is the only analysis-ready data offered from the Landsat program.

2.5.2 Scene Retrieval

Landsat data is owned by the USGS but data can be accessed through online repositories such as EarthExplorer, GloVVis, and the Earth Resource Observation and Science (EROS) Center science-processing architecture (ESPA) on-demand interface. Google Earth Engine (GEE) allows users to download data without any storage infrastructure and to use algorithmic queries to access the data. Amazon Web Services (AWS) offers data access solutions through its Simple Storage Service (S3). Google Earth is the most cited platform for use with Landsat data in conjunction with land use change detection (Hemati et al., 2021).

Landsat images are stored as geographic tagged image file format (GEOTIFF) files. A TIFF or TIF is a Tag Image File Format that supports multiple images within the file which in our case show up as image bands (Shapiro & Stockman, 2001). GEOTIFF files have a geographic tag on them (Sayler, 2022). TIFF files are provided with a lossless or lossy compression, with lossless compression the original image is recovered exactly as it was, which with lossy the pixels cannot be recovered exactly and sometimes

results in data loss (Shapiro & Stockman, 2001). Landsat GEOTIFF files for collection 2 are stored in a lossless format.

2.5.3 Ancillary Data Sources

The Digital Elevation Model (DEM) is produced but the U.S. Geological Survey and is currently used by the MRLC in the production of the NLCD. DEM is used for topographic information (Adhikari et al., 2015) specifically relative elevation, slope, and aspect. This type of topographic information can help distinguish between forest and water classes (Benediktsson et al., 1990). When analyzing coastal habitats, elevation topographic data have seen increases in overall accuracy of up to 31.5% between green spectral features (Letard et al., 2022). The 2001 NLCD included ancillary data to include slope, aspect, and topographic position indexes. The 2016 NLCD included over 30 national-scale ancillary data sets. (Jin et al., 2019)

Sentinel-2 has a near-identical mission of land classification as the Landsat satellites, but is run through the Copernicus Programme by the European Space Agency (ESA) and the European Union (EU). The primary benefit of Sentinel-2 over Landsat is the higher 10m x 10m resolution over Landsat's 30m x 30m resolution. Similar to Landsat, the Sentinel-2 data have significant preprocessing to fix atmospheric, topographic, shadows, cloud cover, and geometric problems (Phiri et al., 2020). Sentinel-2 began in 2015 and does not yet have the multi-decade history that Landsat does, and for this reason, it is not used by the MRLC or the NLCD. While new sensing technologies expand the capabilities of object recognition and classification (Socher et al., 2012) Landsat itself prioritizes offering consistency within its archive over increased sensor strength (Hemati et al., 2021).

2.5.4 Sampling and Validation Techniques

The 1992 NLCD manually reviewed data using aerial imagery and resolved conflicts using ancillary data. The 2006 NLCD models used 30m resolutions, but for the accuracy assessment higher resolution data was used specifically the National Agricultural Imagery Program (NAIP) produced 1-2m resolution data, most likely providing a higher accuracy of the corrected labels. During the design of the 2011 NLCD, it was noted that resource constraints led to a reduction in sampling pixels from 15,000 to 8,000, and the stratification of data was reduced from 10 geographic areas to 2. The geographic areas defined were eastern vs western United States these were included due to the difference in accuracy noted during the 2006 accuracy assessment. The most interesting development out of the 2011 NLCD is the creation of a blind label assessment procedure (Wickham et al., 2017). The 2016 NLCD increased its training samples to 5,000 per class and two models were used, a lite model that focused on land cover types and a full model used for urban and wetland classes. This two-model approach saw improved classification of forest theme classes (Jin et al., 2019).

2.6 Machine Learning Models

Choosing a classifier for the LULC classification requires an understanding of the trade-offs between model types while attempting to break the current 77.5% accuracy rate. From 1992 - 2016 the NLCD thematic assessments showed that the models used for classification were regression tree (CART) algorithms with varying levels of success from 60% to 83% accuracy rates. The 2019 NLCD thematic assessment noted the use of random forest models with a 77.5% accuracy rate on exact matches for both levels of the classification hierarchy and is recorded as the most accurate model assessed under the

strictest evaluation protocol (Wickham et al., 2023). At first glance, the 2019 accuracy appears to be lower (77.5% < 83%); however, the performance difference is methodological, not true degradation. The lower accuracy is in response to the stricter evaluation protocols, such as the decrease in the number of accepted alternative labels in the accuracy assessment by approximately 10%. In 2021, the NLCD adopted an ensemble U-Net classifier; however, it has not gone through a thematic accuracy assessment (Brown, 2024, p. 1).

Table 2.6 - NLCD Model History

Year	Method	Methodology Changes	Accuracy
1992	Tree based algorithm	Leaf on, leaf off	60%
2001	CART	Growing season, topographic	78.8%
2006	CART	Leaf on, leaf off	78%
2011	CART	Leaf on, leaf off	83%
2016	C5.0 Decision Tree	Leaf on, leaf off	82%
2019	Random Forest	Leaf on, leaf off	77.5%
2021	U-Net ensemble	Leaf on, leaf off	Not Yet Published

The hesitation in model upgrades and shifting assessment rules highlights the balance for the NLCD, where both its national visibility and dedication to consistency require methodological caution. Caution can also slow adoption of new techniques, which presents an opportunity to apply a more modern accuracy framework coupled with state of the art architectures to provide greater model accuracy.

2.6.1 Tree based Algorithms

Tree based machine learning algorithms have been used for the majority of the history of the national land cover database and have been well known as one of the most accurate and widely used algorithms for remote sensing classification applications up until at least 2018 (H. Zhao et al., 2018). Historically land classification assessors have selected maximum likelihood, spectral angle mapper, and decision tree based algorithms, produce the highest accuracy over other models. Maximum likelihood classification can perform well when pixel values are normally distributed within each class and have distinct well separated classes. Decision trees create hierarchical decision rules based on the spectral features of each class (Rawat et al., 2024). NLCD in particular is focused on lower resolution imagery with often overlapping classifications, here the widely used decision trees experience overfitting and fail to capture complex global relationships limiting their accuracy for this type of data. More commonly used in the last 10 years are random forest classifiers, which aggregate many decision trees as a single model. Random forest models have effectively overcome the decision tree's limitation of handling high dimensionality and reduces the models risk of overfitting while increasing accuracy (Rawat et al., 2024).

2.6.2 Convolutional Neural Networks and Deep Learning

Deep learning was inspired by the idea that the mammalian brain organizes thought processes with multiple levels of abstraction. They have been applied to fields such as audio recognition, classification tasks, natural language processing (NLP), and computer vision where on average they outperform their traditional model predecessors (L. Zhang et al., 2016). In recent years deep learning has been applied to various fields

within computer vision and remote sensing with great success, so much so that deep neural networks (DNNs) have been adopted by the MRLC for the NLCD. Deep learning techniques can learn feature representations directly from raw pixel values, capturing complex pixel level patterns as well as spatial relationships between neighbors (Rawat et al., 2024).

Deep learning extends the work of single models by allowing for multiple processing layers to learn and represent the data, abstracting intricate structures of large scales of data by mimicking how the brain perceives and understands multimodal information (Voulodimos et al., 2018). These layers can include convolutional, pooling, recurrent cell layers, they can include feeding data forward and backwards, and have memory cells. While the number of possible layer types and combinations is beyond the scope of this research, this praxis highlights the deep-learning model classes particularly well suited to land-cover classification, with particular attention to convolutional and transformer-based vision models. Deep convolution neural networks (CNNs) hold particular promise because they integrate low, mid, and high level features and that these features can be enriched by adding or stacking layers in some cases up to thirty layers (He et al., 2016).

CNNs are designed using deep learning architecture with convolutional operators and excel at computer vision tasks (Kayhan & Gemert, 2020). CNNs are able to automatically extract features for classification by creating overlapping patches of the image for processing called local receptive fields allowing for high accuracy with raw or minimally preprocessed pixels (de Andrade, 2019). One of the first and most influential CNNs for image classification was AlexNet, which upon its creation won the ImageNet

Large Scale Visual Recognition Challenge in 2012 (ILSVRC-2012), outperforming all other methods on the 1.2 million image, 1,000 category dataset by applying rectified linear unit (ReLU) activation, graphical processing unit (GPU) training, an 8 layer architecture, and dropout (Krizhevsky et al., 2017). Other influential CNNs for computer vision consist of LeNet, ZFNet, VGG Net, Google Net, and ResNet. The two most prominent of these —VGG and ResNet— were trained on the ImageNet dataset with millions of training points and offered accuracy significantly beyond competing alternatives (Swapna et al., 2020).

2.6.3 Single and Multi-Pixel Classifications

Tree based algorithms and CNNs classify one label per image, using point of observation values at the single pixel level. These types of models are simple, require lower compute resources, and have good model explainability. However, they lose all information about neighboring pixels which could be of significant use in LULC. The lack of these neighboring pixels also make human evolution of the model or troubleshooting errors difficult. Multipixel classifications or semantic segmentation classifies each pixel within an image as a class, adding complexity and higher computer resources, but solve the downfalls of single pixel classification (J. B. Campbell et al., 2023).

2.6.4 Semantic Segmented Neural Networks

Fully Convolutional Networks (FCNs) were one of the first designs to adapt pixel-based convolutional networks like AlexNet for semantic segmentation, replacing fully connected layers with convolutional ones to enable dense pixel level classification (Long et al., 2015). While FCN's did not surpass the accuracy of state of the art models,

they drastically simplified the data preparation and post processing steps while keeping state of the art accuracy levels. The first semantic segmented models like FCNs included a spatial pooling layer of local features to construct a feature vector descriptor, it was later found that a second-order pooling layer enriched feature interactions capturing pairwise dependencies between local descriptors which increased mean average precision (Carreira et al., 2012).

The most widely used semantic segmentation model is the U-Net architecture, which started from biomedical image segmentation (Ronneberger et al., 2015) but was quickly adopted for land use and land cover as it has shown accuracies of over 90% (Clark et al., 2023). The U-Net architecture adopted FCN's core idea of using only convolution and up sampling layers, however the U-net structure is unique in its addition of multiple skip connections (Ronneberger et al., 2015). Skip connections improve the information flow between layers by breaking symmetries between encoding and decoding paths, allowing local features lost in deep architectures to be retained in between layers(Orhan & Pitkow, 2024).

The team that created the U-Net architecture saw a 10% gain in microscopic cell tracking, winning the IEEE International Symposium on Biomedical Imaging (ISBI) cell tracking challenge in 2015. The link between land use classification and cell tracking in the ISBI challenge comes from the shared reliance on extending model capabilities beyond single pixel classification to multiclass or segmentation classification (Ronneberger et al., 2015). When applied to the detection of palm tree plantations, U-Net was found to be 10% more accurate and 100 times more efficient than AlexNet

(Freudenberg et al., 2019). Adopted by the NLCD in 2023, the current UNET is using 256x256 spatial chips as input, all previous models were pixel-based (Brown, 2024).

2.6.5 Transformers

Transformers evolved from recurrent and convolutional neural network architectures that include encoder and decoder structures, with the most accurate models incorporating attention mechanisms. Unlike the models they evolved from, transformers are built entirely on attention mechanisms, removing the need for recurrence and convolutions altogether (Vaswani et al., 2017). The attention mechanisms allow the capturing of long-term dependencies between elements (S. Khan et al., 2022). Since 2023, transformers have become dominant in many applications of natural language processing (NLP), such as large language models (LLMs) (Burkov, 2025). Even though transformers have seen great success in NLP, in computer vision, they have not experienced the same high level of adoption rates. In computer vision, convolutional architectures remain dominant, but researchers have achieved some success by applying attention mechanisms to convolutional neural networks (CNNs). However, in large-scale image recognition tasks, ResNet like architectures are still the dominant force (Dosovitskiy et al., 2021).

In language models, the strength comes from the relationship between words, while in vision, the relationship is between pixels. For this reason, single-pixel analysis, like those completed in the past using random forest, are no longer suitable. However, in the reverse, if completed like an LLM, there would be a connection between every pixel and every other pixel, which for remote sensing images does not always scale. Instead remote sensing models are applied to local neighborhoods of pixels instead of globally

(Dosovitskiy et al., 2021). Transformers have been adapted specifically for these computer vision problems with the introduction of vision transformers (ViT), which use pure transformers applied directly to image patches to classify images and solving a number of computer vision problems while achieving exceptional performance (Han et al., 2023).

The first vision transformer was the Uniform-scale Vision Transformer (ViT) which showed it was possible to replace convolutional networks with multi-head self-attention applied at a consistent scale to the input image. In image segmentation scenarios like medical and land use classification ViT exceeded the performance of the state-of-the-art models at the time (S. Khan et al., 2022). After ViT, there have been many variants of it, most notably the Data Efficient Image Transformer (DeiT), Shifted Window Transformer (Swin), Tokens-to-Token (T2T), Pyramid Vision Transformer (PVT), and the Conditional Positional Encoding Vision Transformer (CPVT). DeiT was trained on smaller datasets with fewer computer resources while getting equivalent results to ViT. Swin reduces computation by using a window sampling method, shifting the window between layers and a benefit over ViT by capturing fine-grained details with smaller computational loads (M. Khan et al., 2024). DeiT and Swin offer state-of-the-art vision transformer model performance with lower compute requirements.

In 2023 Yao et al. developed an Extended vision transformer (ExVit) that utilized both ViT and cross-modality attention (CMA) to handle long-range correlations among tokenized features in multimodal remote sensing data. The model was only applied to data in Houston Texas and they did not use the NLCD classification system but they achieved a higher overall accuracy than ViT, CCR-Net, or Co-CNN. Some classes did

receive sub-optimal recognition; however, these classes are not present in the NLCD classification system. During the research, they noted the importance of a separable convolution module to extract local details, tokenizing each image tile section, and the use of fusion instead of summary token all contributed to a more accurate classification (Yao et al., 2023).

2.6.6 Unsupervised Classifications

Unsupervised classification excels at discovering new land use categories where they are not already defined but struggles with land use and land cover classification as the natural groupings that arise lay within the spectral classes rather than our defined land use classes. While models can be trained with a hybrid approach to alleviate these class differences, the spectral range between seasons and year over year makes this a difficult task to apply unsupervised classification models (J. Campbell & Wynne, 2011). Because of these limitations the NLCD has historically defaulted to supervised methods as an approach that performs with consistent accuracy for land use problems.

2.6.6 Hybrid Neural Network Transformers

CNNs have dominated hierarchical task-specific feature representation but lack global context. Researchers achieved increased success with deeper networks, but even they started to lose low-level details and suffer from overfitting. Transformers on the other hand excel at global context. Vision transformers like ViT, DeiT, and Swin have problems with fine-grained details because they lack the CNN's ability to operate on overlapping neighbors or spatial inductive bias. The need for both global context and low-level feature extraction has given rise to using CNNs to extract low-level features, then passing them through transformer models to gain global interactions. TransUnet was

among the first to accomplish this using skip-connections to ensure low level features are not lost. TransFuse was accomplished using a BiFusion module with one branch as a shallow CNN and the other as a Transformer allowing them to work in parallel. The shallow CNN not only reduced dependencies on resources but also resulted in retaining richer local information, and because of its simplicity, almost any off-the-shelf convolution network can be used. These models achieved up to a 10% gain in accuracy over ResUnet and U-net over 5 test sets (Y. Zhang et al., 2021).

Since TransUnet was designed a number of hybrid models have come forward with their own unique benefits. Guided lightweight multiscale network (GLIS) used a convolutional network and Swin saw reduced compute resources with industry-leading accuracy (Yazıcı et al., 2024). PHTrans utilized the U-shaped encoder-decoder for early feature extraction with transformer and convolutional layers in deeper stages, achieving high accuracy but also increased model complexity and resource requirements (W. Liu et al., 2022). Other models comparable in accuracy and varying complexity trade-offs like FIF-UNet (Gou et al., 2024), Hiformer (Heidari et al., 2023), and STransU2Net (G. Liu et al., 2024) have been applied to imagery problems with great success. While STransU2Net boasts a notable ~5% gain in accuracy over Unet models and a ~3% accuracy gain over SwinUNet (G. Liu et al., 2024), making Trans2Unet a state-of-the-art leader.

2.7 Conclusion

The National Land Cover Database (NLCD) is an indispensable product used in numerous applications of high importance to the monitoring and solutioning of our evolving world. Spectral data, seasonal timing, and modeling tools have created

bottlenecks moving accuracy forward by only a few percentage points since 2001. The review in §2.2-§2.6 highlights three decades worth of advancements in computer vision, remote sensing technology, feature engineering, and machine learning (ML) techniques that have created the following gaps and opportunities.

Table 2.7 - Literature Review Key Findings

Key Findings from Literature	Research Gaps and Opportunities
§2.2 The Multi Resolution Land Characteristics Consortium and the National Land Cover Database	
• Authoritative U.S. land use classification product. • Classifications are Anderson level 1 and level 2 based.	• No published accuracy for U-Net. • Methodological Consistency has led to slow innovations
§2.3 Spectral Features	
• Surface-reflectance Landsat bands drive class separability yet problems exist in mixed land covers.	• Limited spectral indices usage. • Spectral derivatives assist in higher accuracy rates.
§2.4 Temporal Consistency and Phenology	
• Seasonal confusion persists. • CNN / RNN models effective for sparse data.	• Current temporal window is rigid. • Limited research in hybrid temporal transformers.
§2.5 Scene Acquisition and Archiving	
• Cloud-free coverage is problematic. • DEM: slope, aspect, and topography increases accuracy.	• GEE facilitates data integration and retreival. • Adaptive sampling schemes and probabilistic accuracy assessments should be integrated.
§2.6 Machine Learning Models	
• CART/C5.0 (60-80% accuracy), Random Forest (77.5% accuracy), U-Net (recent adoption, no accuracy). Transformers and Hybrids (>90% in regional testing)	• National benchmarking of transformers and hybrid models absent. • Computation/Storage constraints for national benchmarks need quantification

Persistent challenges include limited exploitation of spectral indices and derivatives, class specific temporal windows and rigidity. Current state-of-the-art

architectures lack national benchmarking, and understanding of the computational and storage requirements to do so is limited. Random forest is a solid baseline model, it has been recognized for its contribution and application to remote sensing and specifically its use within the NLCD for many years. U-Net boasts high results among convolutional network architectures as evident by its recent adoption by the MRLC. Vision transformers have seen even greater accuracy results than U-Net but only in limited applications with less variability than the NLCD. CNN-transformer hybrids show promising accuracy scores under limited and regional applications. This praxis should prioritize the national-scale benchmarking of advanced models, explore temporal adaptations, the integration of spectral indicis and derivatives while monitoring computational resources. These efforts offer national land cover classification advancements in the accuracy and reliability of the data used to support its environmental, policy, and land management practices.

Chapter 3 – Methodology

3.1 Introduction

This chapter outlines the methodology used to evaluate possible improvements to land use and land cover (LULC) classification accuracy beyond the National Land Cover Database's (NLCD) 77.5% benchmark (Wickham et al., 2023) to at least 90%. This praxis aimed to identify state-of-the-art machine learning models for land use classification. The model architecture comparisons (RQ1-RQ3 and H1-H3) focus on identifying the most effective machine learning architectures for LULC classification. These include evaluations of deep neural networks, vision transformers, and hybrid CNN+Transformer architectures to determine their accuracies over Random Forest and U-Net models. Temporal context for accuracy gains (RQ4 and H4) aimed to determine a method to create temporally rich data and validate that it enhances model accuracy. The chapter begins by detailing the data sources, which include surface reflectance imagery from Landsat 8, NLCD land use and confidence labels, ancillary data from the Digital Elevation Model (DEM), Moderate Resolution Imaging Spectroradiometer (MODIS), and the Environmental Protection Agency's (EPA) Ecological Zones. Subsequent sections discuss the application of different spectral derivatives and temporal patterns to offer increased insights into difficult-to-classify NLCD classes. To ensure that any performance differences are due to the models rather than the data ingestion, preprocessing, feature extraction, or training and validation splits, a single end-to-end pipeline is used. This is intended to eliminate data-related biases, imbalances, or leakages that may occur so that the Random Forest, U-Net, transformer, and hybrid models all

learn from exactly the same, consistently prepared data. Lastly, the model evaluation metrics are used to determine the outcomes of the research questions and hypotheses.

3.1.1 Challenges

The literature review (Table 2.2) summarizes gaps and opportunities, identifying several challenges that shaped the design choices of the methodology outlined in this chapter. These challenges highlight roadblocks, inconsistencies, and ambiguities that reduce classification accuracy and reproducibility. This praxis has listed each challenge in Table 3.1.1 below and assigned a shorthand label (C-X) and a short description for consistent reference in the rest of the chapter.

Table 3.1.1 - NLCD Classifier Challenges

C1	**Spectral Confusion**: Classes with similar remote sensing spectral values are easy to classify incorrectly (i.e., Barren land vs Pasture/Hay, Woody Wetlands vs Deciduous Forest, and Developed High Intensity vs Developed Medium Intensity).
C2	**Mixed Pixels**: A 30-meter pixel resolution can include two classes that must be represented in a single pixel value (i.e., Developed Low Intensity sharing a 30-meter pixel space with an Evergreen Forest).
C3	**Phenological Variation**: Seasonal vegetation cycles change spectral signatures, creating a class shift if timing is not accounted for (i.e., cultivated crops in growth or harvest cycles, or deciduous forest during winter or summer months).
C4	**Obscurants**: Atmospheric effects can cause corrupt remote sensing recordings (i.e., clouds, shadows of clouds, snow, and ice).
C5	**Radiometric Saturation and Noise**: Radiometric disturbances can max out remote sensor values.
C6	**Class Imbalance**: Rare classes, such as snow or urban development bias models toward dominant types.
C7	**Continental Scale Data Volume**: Overall imagery size for the entire continental United States (CONUS) is impractical because of the size of approximately 160,000 pixels by 105,000 pixels per layer. Larger than any single download will allow for all projected data.
C8	**Cross-Source Alignment**: Landsat, NLCD, DEM, and MODIS differ in recording geometry. Even when reprojecting all the sources into a single method, this can cause null values along the edge of the images.

C9	**Regional Domain Shift**: Land cover classes have different spectral signatures across the U.S. Even something as simple as a 'forest' will have a vastly different signature in the Pacific Northwest than in Appalachia.

3.1.2 Methodology Workflow

The methodology is ultimately a foundational step that supports the overall goal of a 12.5% improvement in NLCD land use classification and is structured into five main stages, as shown in Figure 3.1.2 below: First, the dataset is defined, acquired, and structured (§3.2). Next, the data preprocessing and exploratory analysis ensure data quality and identify foundational patterns (§3.3). Then, the feature engineering creates spectral, textural, and temporal features (§3.4). That data is supplied to the model training and optimization pipeline to refine machine learning models (§3.4). Lastly, model evaluation (§3.6) checks performance, providing a clear metric-based approach to the comparison of models relative to the NLCD baseline figures.

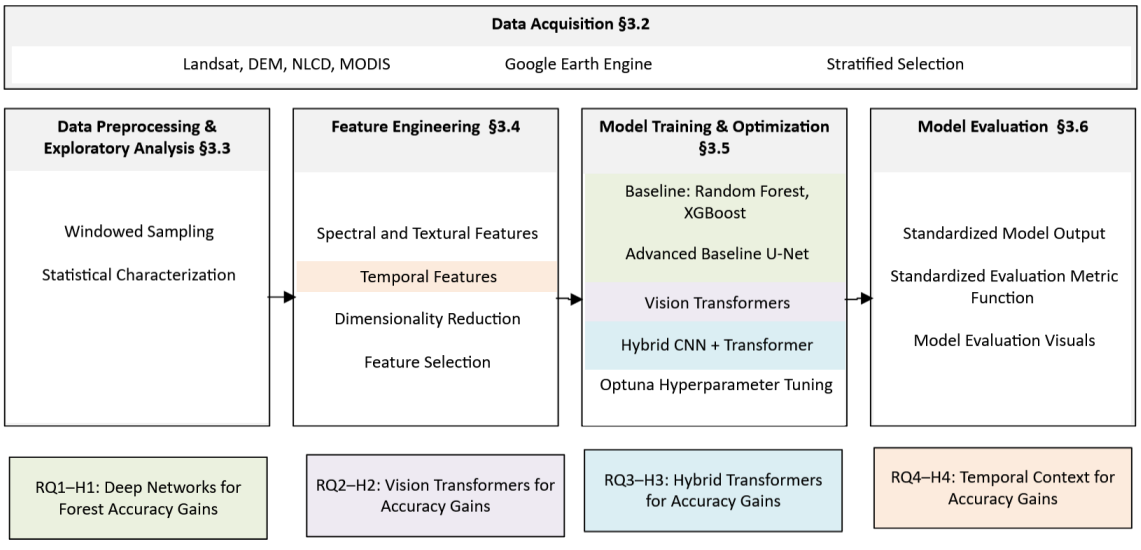

Figure 3.1.2 – Methodology Overview

3.2 Data Acquisition and Structure

Google Earth Engine (GEE) is used to structure and download all source data. GEE is a cloud-based platform that offers geospatial server-side computing. GEE provides the download of the following data sources. Landsat images (layers) are the primary source of satellite remote sensing training data for this praxis. The NLCD land use classification is the Multi-Resolution Land Characteristics (MRLC) Consortium's authoritative land use data source. While the NLCD supplies the pixel-level labels and confidence estimates, it lacks the ancillary data necessary to match the variables historically used and outlined by the MRLC (Brown, 2024). These missing data points, such as elevation, phenology, and regional characteristics, were collected from the data sources that follow. DEM data is added to the training data to increase model accuracy. MODIS phenological data are used to create phenological date ranges for Landsat. EPA ecological zones are used to create region-aware features. NLCD confidence score layers are used as data quality filters and are imported from the MRLC into GEE. Details of these five data sources are outlined in §3.2.1.

The data is split into 2,000 pixel by 2,000 pixel squares of CONUS in order to handle the continental-scale data volume challenge (C7). Each point has a European Petroleum Survey Group (EPSG) Universal Transverse Mercator (UTM) code, which sets a standard display for all different data sources to eliminate cross-source alignment challenges (C8). The MODIS phenological dates are used to gather surface reflectance, temperature, and quality assurance (QA) layers for both leaf-on (summer) and leaf-off (winter) periods. Phenological variation (C3) and temporal context for accuracy gains

(RQ4-H4) necessitate the need for these MODIS dates. The process of gathering split temporal Landsat data, reprojecting them to the same EPSG code, and merging all the data sources is handled by GEE and outlined in §3.2.7. The output described in §3.2.7 is the downloading of all Tagged Image File Format (TIFF) data as the inputs for the process described in §3.3. This methodology is outlined in Figure 3.2 below.

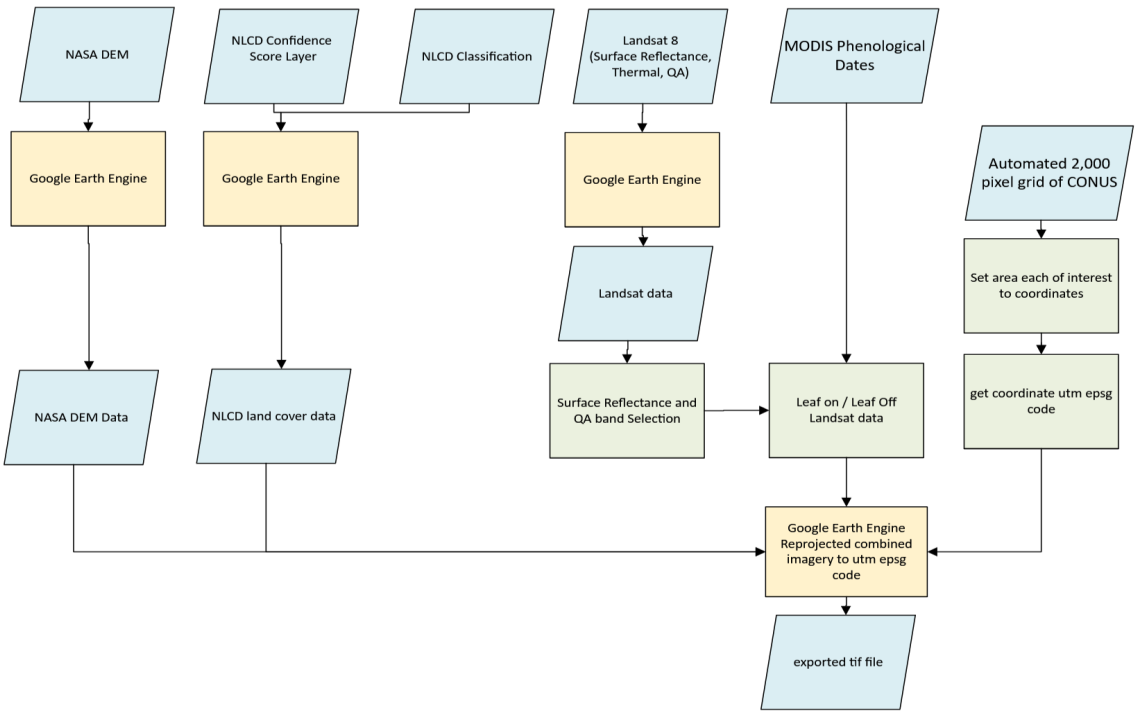

Figure 3.2 – Raw Data Creation

3.2.1 Data Source: Landsat

Landsat data provide the core of the input training data for this praxis, supporting model development for all research questions and hypothesis. Landsat captures information from many different sensors and offers a suite of data for use. These data are

readily available from GEE. This praxis utilizes data from the surface reflectance, thermal radiation, and data quality bands. Each measurement taken by Landsat is stored in separate image layers (bands) within a Geographic Tagged Image File Format (Geo TIFF) in 30-meter per pixel resolutions. This praxis utilizes the bands listed in Table 3.2.1 for each area of interest (AOI) and time period (defined below in the sections §3.2.5 and §3.2.6, respectively).

Table 3.2.1 - Landsat Bands

Surface Reflectance Bands	Coastal aerosol, Blue, Green, Red, Near Infrared, Shortwave Infrared 1, Shortwave Infrared 2
Thermal Band	Thermal Infrared 1
Data Quality Bands	- The Quality Assessment Pixel-level bands (QA_PIXEL) is a bit level combination that defines certain quality conditions (snow, clouds, cloud type, and confidence of cloud) - Radiometric Saturation (QA_RADSAT) is a representation of which sensor bands were saturated during data capture, resulting in unusable data. - Aerosol (QA_AEROSOL) is a bit representation of the interpolation or raw value delivered for water identification. - Surface Temperature (ST_QA) is a bit representation of the confidence of the surface temperature retrieval. - Cloud Distance (ST_CDIST) is a measurement in meters of how far the recorded pixels are from a detected cloud. - Surface Emissivity (ST_EMIS) is an estimate of how efficiently the surface emits thermal radiation.

3.2.2 Data Source: NLCD

The NLCD land use classification data serves as the ground truth labels in support of all research questions and hypothesis. The NLCD is the authoritative land use/ land cover classification dataset for the United States, produced by the MRLC, and is readily available from GEE. The NLCD land cover confidence is downloaded from the MRLC

and imported into GEE to utilize GEE's geospatial alignment processing. This praxis uses two of the outputs from the NLCD, as shown in Table 3.2.2 below.

Table 3.2.2 – NLCD Data

Land Cover Classification	Depicts predominant land cover classes and is used as the ground truth labels. Land cover classes are hierarchical pixel-level categories with 8 level 1 classes and 16 level 2 classes. The full class list can be found in Appendix Table A4. The classification is set on a 30-meter per pixel resolution similar to Landsat imagery.
Land Cover Confidence	Depicts the probability value for the land cover class identified by the MRLC's classification method at a 30-meter resolution. The land cover confidence is used to filter out areas where the current NLCD models are unsure of the land classification type.

3.2.3 Data Source: DEM

This praxis utilizes topographic data from the National Aeronautics and Space Administration (NASA) Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM). DEM Data is included to enhance the accuracy of the models by adding topographic data to the Landsat images. While other elevation data sources are available, NASA's SRTM provides consistent global coverage at a 30m resolution, aligning with the NLCD's resolution and the consistency goals of the MRLC. Data includes the layers outlined in Table 3.2.3 below:

Table 3.2.3 - DEM Data

Elevation	Surface Elevation in meters.
Slope	Ground steepness in degrees.
Aspect	Downslope compass direction in degrees (0-360).
Hill Shade	Shade illumination (0-255)

3.2.4 Data Source: MODIS

This praxis utilizes temporal phenological data from the MODIS MCD12Q2 dataset. This phenological data is used to answer temporal context for accuracy gains (RQ4 and H4). These data also are utilized in the model architecture comparisons (RQ1-RQ3 and H1-H3), aiding in the phenological variation challenge (C3). These data contain the key dates of phenological change for the United States at 500-meter resolution. The dates at which vegetation starts, peaks, and ends can be drastically different based on its geological position. These data provide the dates in Table 3.2.4 below for each area of interest in the source data.

Table 3.2.4 – MODIS Data

Green Up	The first growth of the year.
Mid Green Up	The halfway point between first growth and peak growth.
Mid Green Down	The halfway point between peak growth and dormancy.
Dormancy	The point at which vegetation remains inactive until the next green-up cycle.

3.2.5 Data Source: EPA Ecological Regions

This praxis utilizes geographic regional zones from the EPA ecological region (ecoregion) data. These geographic data are used as a solution to the regional data shift challenge (C9). Ecological regions are delineations of discrete systems that comprise of geological, landform, soil, vegetative, climatic, wildlife, water, and human factors (Pisanty et al., 1997). The data for these regions is downloaded from the EPA in the form of a 30 MB shape file, which contains the geospatial geometries that are overlaid with AOI coordinates for ecoregion retrieval.

Table 3.2.5 - EPA Ecological Regions

Level 1	Contains 15 ecoregions, 8 of which are within CONUS. Examples include: Northwestern Forested Mountains, Great Plains, Mediterranean California, Tropical Humid Forest, and Eastern Temperate Forests.
Level 2	Contains 52 ecoregions, 21 of which are within CONUS. Examples include: Marine West Coast Forest, Central USA Plains, and Everglades.

3.2.5 Geographic Tiling

To handle the continental-scale data volume challenge (C7), this praxis utilizes a grid system over CONUS comprised of 2,408 AOIs. In its entirety, the download of data surpasses the maximum export size allowed from GEE. Using small AOI areas requires increased processing time by GEE, estimated at 55 days. Therefore, a 2048x2048 pixel resolution was set because it optimized processing time and allowable download size. Exporting in 2048x2048 resolution results in an 80 x 53 grid with a 48 pixel overlap to remove issues related to cross-source alignment (C8). With the latitude and longitudinal points saved in a comma-separated values (CSV) file, the entire U.S. is covered by this sampling method. The aforementioned CSV is utilized in the data integration process outlined in §3.2.7.

3.2.6 Temporal Features

Phenologically temporal features help the distinction between certain classes, such as deciduous forest versus evergreen forest, grassland versus pasture, and shrub versus cultivated crops. These phenological features are included in this praxis to address the temporal context for accuracy gains (RQ4 and H4) as well as to offer a solution to the phenological variation challenge (C3) in the model architecture comparisons (RQ1-RQ3 and H1-H3). The accuracy of land use models can be increased by adding a temporal

factor, as the remote sensing values during winter and summer show dramatic differences. For example, the spectral values of deciduous and evergreen forests from winter to summer months. While summer months are almost identical, winter values are dramatically different. Because of these differences in phenology, it's become customary to use a leaf-on / leaf-off temporal approach. In particular, taking images of both peak season and dormmate season to increase classification accuracy.

The current NLCD method is to create a synthetic surface reflectance value for November 15th of each year via regression models and deep learning models for their leaf-off dates. This praxis instead utilizes the MODIS data described in §3.2.4. For each geographic tile defined in §3.2.5 the MODIS green up, mid green up, mid green down, and dormancy dates are queried as seen in Table 3.2.3 below. These dates are used in the processes outlined in §3.2.7 to query GEE for two phenological periods:

1. Leaf on: from mid green up to mid green down, offering coverage of the peak growing season
2. Leaf off: from dormancy to green up, covering the dormancy plant growth periods.

Table 3.2.3 - MODIS Phenological Dates Sample

longitude	latitude	midgreendown	midgreenup	greenup	dormancy
-125.1905198	48.49475094	2019-09-04	2019-04-28	2019-04-01	2019-10-06
-117.9600569	49.27553265583022	2019-09-12	2019-05-25	2019-05-01	2019-10-09
-93.38521067	48.762565660749175	2019-09-17	2019-06-02	2019-05-07	2019-10-13

While leaf on periods typically contain clear Landsat imagery, leaf off periods are often cloudier and therefore it is difficult to obtain unobstructed scenes. This praxis uses

the top 20 clearest images for each period, based on the QA_Pixel defined in §3.2.1. From these 20 images a single composite is produced by choosing the highest quality pixel for each pixel in the composite. This produces a single, clearest pixel image. However, this method alone was not enough to ensure a completely unobstructed scene. When no quality pixel is available for the current year's period, the previous years 20 clearest images are incorporated. The result is a single composite image with the clearest single pixel values for 40 different images, prioritizing the current year. This strategy of creation composites that include even the previous year's values to inform the current year's land use classification is a tactic currently utilized by the MRLC in the production of the NLCD (Brown, 2024, p. 1).

3.2.7 Data Integration

GEE is utilized to bring the data described in §3.2.1-§3.2.6 into a single TIFF file to address the cross-source alignment (C8). GEE completes the spatial and temporal processing for this praxis. Most importantly to this praxis is spatially aligning the different data sources. All layers from Landsat, NLCD, DEM, including temporal layers are reprojected using the coordinate reference system (CRS) EPGS code. Without the CRS and EPGS codes the different layers such as a river in Landsat imagery and the land classification for the river may be a few pixels away from each other. GEE also computes temporal dates and the quality composites defined in §3.2.6 which ultimately builds our complete raw data set.

Google Colab was used to programmatically download each area of interest from the CSV's latitude and longitude points identified in §3.2.5 during the geographic tiling. A Python script was written to loop through the CSV to initiate GEE processing of each

AOI and layers from each are combined into a single TIFF file. Google Collab is connected to GEE and Google Drive storage systems saving the TIF files in a convenient location for local download. Out of the 2,408 AOIs to export, 32 AOIs resulted in GEE processing errors that could be a result of many different computational and filtering factors. The 32 missing AOIs span from southern California to Texas and account for 134 million pixels compared to the 9.9 billion that remain. The 2,376 remaining files are downloaded locally and ready for data preprocessing and exploratory analysis.

3.3 Data Processing and Exploratory Analysis

This praxis utilizes exploratory analysis to reveal important insights and data preprocessing to prepare the data for feature engineering. While data preprocessing typically focuses on data cleansing and transformations, for this praxis, the focus is placed on finding the clean and correctly transformed data within the dataset. The data preprocessing for this praxis begins with the 2,376 AOIs, which consist of 2048x2048 TIFF files output from the process described in §3.2.4. The goal is to systematically create high-quality training data, reducing the large AOIs into windowed 64x64 patches outlined in Figure 3.3 below. This is detailed in §3.3.1 using adaptive strides and mean confidence scores to meet minimum class representation. These windows are then filtered to exclude bands with missing data, including anomalies, and stricter quality requirements are applied to high representation classes in §3.3.2. Lastly, in §3.3.3, all the saved scenes are grouped by their original AOI and assigned to a train, validation, or test dataset to prevent data leakage from overlapping windows across classes. The data preprocessing takes place in Visual Studio Code on a local desktop with an Intel i7-14700k 3.4 GHz CPU, 128 GB of 4800MHz Ram, Nvidia 4070 super 12 GB GPU, 2TB

M2 storage, and a 24TB HDD storage. The output from this section is train, validation, and test folders, each filled with its assigned TIFFs and ready for feature engineering in §3.4.

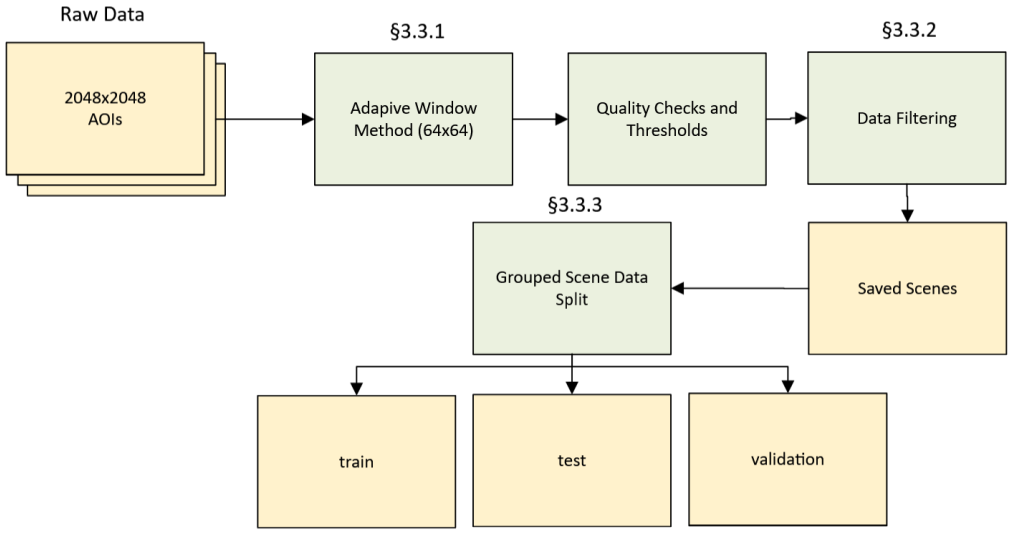

Figure 3.3 – Windowed Sampling Method

3.3.1 Adaptive Window Method

The adaptive window method was developed to address the class imbalance challenge (C6). This method is inspired by the window-based tiling methods in segmentation models such as the shifted window (Swin). It allows for low-representation classes, such as snow and high-intensity development, to be oversampled via overlapping windows in a method. The process is completed by a Python script that crawls through each 2048x2048 TIFF and saves 64x64 TIFF images. The window method starts by

defining a per-class threshold of 50,000. This praxis assumes a 70/15/15 train/validation/test split and a reduction of overall pixels due to quality, missing bands, and AOI prioritization, so that the final training set contains at least 10,000 samples. This praxis uses U-Net and Transformer segmentation models. The sample count is set to 10,000 as a pragmatic midpoint within the ranges reported in prior studies. Looping through with each class, it starts at the highest stride and confidence window. Stride describes how many pixels the window moves between each read, 64 pixels meaning no overlap, while 32 is a 50% overlap. The confidence threshold is the mean confidence score for the window, starting with 80. Using the mean ensures that hard-to-interpret pixels are not ignored while ensuring the model isn't trained on mostly noise.

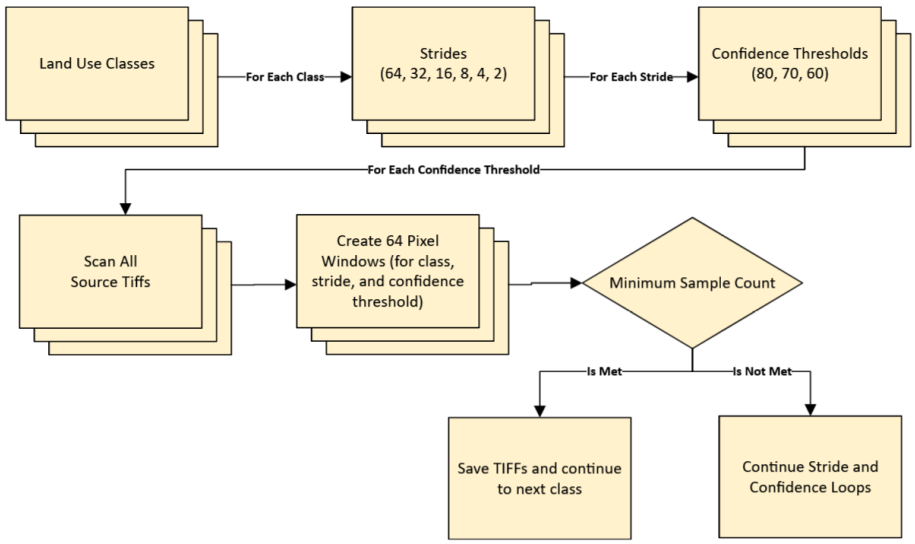

Figure 3.3.1 - Adaptive Window Method

3.3.2 Data Filtering

Data filtering is conducted to identify and remove scenes that include data that may erode model performance. The windows saved during the processes outlined in §3.2.1 that contain null pixel values for land use, land confidence, Landsat, or DEM values are also removed. Dominant classes are thinned by progressively increasing the confidence requirement of high representation classes with a goal of 6.25% per class representation. The Landsat QA_Pixel values are used to remove windows with obscurants (C4). RAD_SAT values are used to remove windows with radiometric saturation and noise (C5). Lastly, a 64 pixel border around each AOI (2048 pixel image) is ignored to create a buffer zone, reducing the chance of data leakage. Throughout this process, statistics of land cover confidence and class cooccurrence are taken into account to minimize negative impacts on low-representation classes.

3.3.3 Grouped Scene Data Split

The clean data from the methodology in §3.3.2 are grouped together based on their AOI and sorted in descending order by the number of pixels per class. Starting with the rarest of classes (snow) AOIs are assigned to train, validation, and test datasets. As all classes and AOIs are processed together, this helps ensure proper stratification of balanced classes at the AOI level. The result of the data processing and exploratory analysis is three folders: train, validation, and test, each filled with thousands of 64x64 pixel TIFF images. Each data folder is maximized for label confidence, enhanced for remote sensing clarity of image, void of data leakage issues, and ready for feature engineering.

3.4 Feature Engineering and Selection

Feature engineering for this praxis involves developing one-hot encoded features, spectral indices, and textural descriptors. Feature selection is conducted to identify the most relevant features for the models used in this praxis. Creating and selecting highly relevant features trains models for the model architecture comparisons (RQ1-RQ3 and H1-H3) on useful signals and improves class separability. One-hot encoded features are created in the processes outlined in §3.4.1 from EPA Ecoregions and Landsat QA bit-packed bands (QA_PIXEL, QA_RADSAT, and QA_AEROSOL). Spectral indices (X. Zhao et al., 2022) and textural features (S. Liu & Gu, 2017) are widely used methods of increasing classification signals from raw remote sensing data. Overall, 134 features were created: 24 spectral, 50 textural, and 60 QA features. All features are created from both leaf-on and leaf-off imagery, creating temporally distinct features. Relevant features such as surface reflectance undergo standardization for robust model performance. After all the features have been created, a sample of 100,000 pixels from each class is randomly selected to create a dataset for feature selection. Principal component analysis is conducted to determine the minimum number of principal components required to capture 95% of the explained variance. After this, random forest and XGBoost models are used to compute overall and class-specific feature importance (Gini and SHAP) to highlight the importance of specific features. The correlation matrix assists in removing duplicative information that can complicate the models. At this point, the final feature set is selected and applied to the entire train, validation, and test with identical data transformations. The output from this section is the final dataset ready for model training and optimization.

3.4.1 One-Hot Encodings

Many categorical variables exist within and are added to the data from secondary data sources. Of primary concern is the quality assurance bands provided by Landsat that give sensor-specific quality metrics for the provided Landsat remote sensing imagery used by this praxis. Secondary, is the EPA ecoregion zone 2 which is used to provide regional-specific information about the ecology of the provided imagery to handle regional domain shift (C9).

As covered in §3.2.5 the EPA ecological region data is provided in a shapefile. The file is loaded into Python, and each TIFF's coordinates are queried against the shapefile, which returns the region that the TIFF is located in. These regions are then one-hot encoded into 21 new bands, each representing a specific ecological region within CONUS.

Three of the quality assurance bands delivered with Landsat imagery are bit-packed values that need to be decoded for model use. Having these QA values decoded helps the models understand where raw Landsat values may not deliver consistent information and possibly degrade model performance. QA_PIXEL, QA_AEROSOL, and QA_RADSAT are bit-packed layers (C. Barnes, 2024) decoded according to the Landsat Collection 2 Level-2 Science Product Guide, which contains the bit-mapping tables for these three bands. Each unpacked value is then transformed into pixel-level, creating 58 one-hot encoded bands are created and tested for importance.

QA_PIXEL is converted into 22 new bands, 11 each for both leaf-on and leaf-off periods. These are created for various clouds, snow, or water vapor conditions. Please see Table A-1 in the Appendix for a full breakdown of all the QA Values.

QA_RADSAT is decoded into 20 new bands, 10 each for both leaf-on and leaf-off periods. This band is bitpacked to show the radiometric saturation for each reflectance band and the temperature band. For example, the first bit is either a 0 for no saturation or a 1 for saturated data specifically for band 1. Please see Table A-2 in the Appendix for a full breakdown of all the QA values.

Lastly, the QA_AEROSOL is decoded into 16 new bands, eight each for both leaf-on and leaf-off periods. This band is bit-packed to show if the pixel has recorded aerosol in the sensor readings, which type, and scores that lead to a confidence in that reading. For example, bits 6-7 show at which level the aerosol was detected (low, medium, high). Please see Table A-3 in the Appendix for a full breakdown of all the QA values.

3.4.2 Spectral Indices

Spectral indices are applied to enhance the specific characteristics of land cover classes. This improved information assists models in land use classification, aligning with the model architecture comparisons (RQ1-RQ3 and H1-H3) by transforming surface reflectance bands to highlight certain features or patterns (Huang et al., 2021). These indices combine reflectance values from Landsat remote sensing images to create new derived index bands that emphasize critical surface characteristics, such as water content, soil exposure, and vegetation health. The creation of these new indices improves class separability and can correlate more directly with the physical properties of a land cover class than raw reflectance values. Some of the most published indices used in this praxis, a use case example, and its calculations are as follows:

Table 3.4.2a - Symbols and Definitions for Index Calculations

Symbol	Meaning (per pixel)	Landsat band
ρ_B	Blue reflectance (unitless, surface reflectance)	B2
ρ_G	Green reflectance	B3
ρ_R	Red reflectance	B4
ρ_{NIR}	Near-infrared reflectance	B5
ρ_{SWIR1}	Shortwave infrared 1 reflectance	B6
ρ_{SWIR2}	Shortwave infrared 2 reflectance	B7

Tasseled Cap – Using all image bands this shows surface brightness, greenness, and wetness. First used in the identification of quarry locations, the Tasseled Cap is now used for many applications (H. Zhao et al., 2018). Each set of coefficients is used to create new a weighted sum of the original bands for brightness, greenness and wetness. The weights (F) used are located in Table 3.4.2b below, based on the tasseled cap transformations (TCT) by Baig et al., which identified these weights from thousands of quality pixels. (Baig et al., 2014)

$$\boldsymbol{TC_X = \rho_B * F_B + \rho_G * F_G + \rho_R * F_R + \rho_{NIR} * F_{NIR} + \rho_{SWIR1} * F_{SWIR1} + \rho_{SWIR2} * F_{SWIR2}} \quad (1)$$

Table 3.4.2b - Tasseled Cap Weights

TCT	Blue	Green	Red	NIR	SWIR1	SWIR2
Brightness	0.3029	0.2786	0.4733	0.5599	0.508	0.1872

Greenness	-0.2941	-0.243	-0.5424	0.7276	0.0713	-0.1608
Wetness	0.1511	0.1973	0.3283	0.3407	-0.7117	-0.4559

Normalized Difference Vegetation Index (NDVI) – Used to calculate vegetation status, growth, or the removal of vegetation. Used in the identification of quarry locations (H. Zhao et al., 2018).

$$\boldsymbol{NDVI = \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R}} \quad (2)$$

Enhanced Vegetation Index (EVI) – This is an improved method of NDVI vegetation index that corrects for atmospheric and soil influences (X. Zhang et al., 2012).

$$\boldsymbol{EVI = 2.5 \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + 6\rho_R - 7.5\rho_B + 1}} \quad (3)$$

Chlorophyll Index Green (CI Green) – Uses green edge reflectance to identify chlorophyll (Tanaka et al., 2015).

$$\boldsymbol{CIGreen = \frac{\rho_{NIR}}{\rho_G} - 1} \quad (4)$$

Normalized Difference Water Index (NDWI) – Used to indicate surface moisture. Used to assist in quarry identification via exclusion based on water bodies (H. Zhao et al., 2018)

$$\boldsymbol{NDWI = \frac{\rho_G - \rho_{NIR}}{\rho_G + \rho_{NIR}}} \quad (5)$$

Normalized Difference Snow Index (NDSI) – An index for snow identification, classified as snow if NDSI > 0.4, $\rho_G > 0.10$, $\rho_{SPIR} > 0.11$, otherwise "not snow" (Wunderle et al., 2016).

$$\boldsymbol{NDSI = \frac{\rho_G - \rho_{SWIR1}}{\rho_G + \rho_{SWIR1}}} \quad (6)$$

Normalized Difference Building Index (NDBI) – Used to detect bult-up surfaces such as cement or roads. Used in quarry identification to reduce confusion between quarries and urban development (H. Zhao et al., 2018).

$$\boldsymbol{NDBI = \frac{\rho_{SWIR1} - \rho_{NIR}}{\rho_{SWIR1} + \rho_{NIR}}} \quad (7)$$

Modified Soil-Adjusted Vegetation Index (MSAVI) – Uses adjusted soil factors to minimize soil-induced noise, specifically helps with vegetation recognition in arid and semi-arid regions (Yang et al., 2022).

$$\boldsymbol{MSAVI = \frac{2\,x\,\rho_{NIR} + 1 - \sqrt{(2\,X\,\rho_{NIR} + 1)^2 - 8\,x\,(\rho_{NIR} - \rho_R)}}{2}} \quad (8)$$

Normalized Difference Moisture Index (NDMI) – Used to measure moisture levels within vegetation, signaling wetlands or droughts (Yang et al., 2022).

$$\boldsymbol{NDMI} = \frac{\boldsymbol{\rho_{NIR} - \rho_{SWIR1}}}{\boldsymbol{\rho_{NIR} + \rho_{SWIR1}}} \tag{9}$$

3.4.3 Textural Features

Textural features are created to capture spatial patterns that spectral values alone do not capture. These features can help map how pixel values vary across space, supporting more consistent accuracy (Sim et al., 2024) across the various models used to investigate the model architecture comparisons (RQ1-RQ3 and H1-H3). Gray Level Co-Occurrence Matrix (GLCM) is a spatial feature extraction approach that captures these spatial attributes, such as surface roughness, uniformity, and complexity. Crops, for example, generally produce highly consistent repeating textures rather than a natural wood line or pasture. GLCM is a textural measure to detect sparse and large settlements (S. Liu & Gu, 2017) and is best applied to a single band at a time. The following GLCM measures are applied at the band level, resulting in 10 new bands: 5 for each of the leaf-on and leaf-off periods, as per the following list.

Contrast – Measures the variation of pixel values, specifically helps highlight sharp changes in texture.

$$\boldsymbol{CON} = \sum_{\boldsymbol{k}} \boldsymbol{k^2} \left| \sum_{\boldsymbol{i}} \sum_{\substack{\boldsymbol{j} \\ |\boldsymbol{i-j}|=\boldsymbol{k}}} \boldsymbol{p(i,j)} \right| \tag{10}$$

Dissimilarity – Measures the difference between pixel neighbors within a neighborhood, ultimately describing the coarseness of an image.

$$\boldsymbol{Disim} = \sum_i \sum_j |\boldsymbol{i} - \boldsymbol{j}| \cdot \boldsymbol{p}(\boldsymbol{i},\boldsymbol{j}) \quad (11)$$

Homogeneity (Inverse Difference Moment) – Measures the similarity between pixels, resulting in smooth textures getting high scores.

$$\boldsymbol{IDM} = \sum_i \sum_j [\, \mathbf{1} / (\mathbf{1} + |\boldsymbol{i} - \boldsymbol{j}|^2) \,] \cdot \boldsymbol{p}(\boldsymbol{i},\boldsymbol{j}) \quad (12)$$

Energy (Angular Second Moment) – Measures the uniformity or repetition in textural patterns, low scores indicate randomness and high scores indicate consistent regions.

$$\boldsymbol{ASM} = \sum_i \sum_j \{\boldsymbol{p}(\boldsymbol{i},\boldsymbol{j})\}^2 \quad (13)$$

Correlation – Measures how linearly related pixel values are for any given direction, helps detect structural properties.

$$\boldsymbol{COR} = \sum_i \sum_j ((\boldsymbol{i} \cdot \boldsymbol{j}) \cdot \boldsymbol{p}(\boldsymbol{i},\boldsymbol{j}) - \boldsymbol{\mu}_x \cdot \boldsymbol{\mu}_y) / (\boldsymbol{\sigma}_x \cdot \boldsymbol{\sigma}_y) \quad (14)$$

3.4.4 Standardization

In this praxis, multiple types of standardization are utilized for optimal model performance. Various data inputs are consistently scaled, such as surface temperature bands (ST), surface reflectance (SR), Tasseled Cap indices, and GLCM features. By using a robust scaler for each of these features, the standardization is less sensitive to the

distribution native to remote sensing fields. (Röchner et al., 2025). This helps the data understand outlier scores, aligning with this praxis' methodology surrounding consistency and robustness of the input data.

3.4.5 Feature Selection

Feature selection for this praxis is conducted to reduce the total number of features from 210 to a smaller subset that focuses on the reduction of redundancy and meaningfulness of spectral representation between land use classes. This feature selection process assists in all research questions and hypotheses, as it affects both the model development and temporal representation during development. The process employs a multi-step analysis that focuses on dimensionality reduction using Principal Component Analysis (PCA), Shapley Additive Explanations (SHAP) model-based explainability, Gini ensemble-based importance using random forest, and a correlation matrix. Lastly, to ensure feature selection accuracy, features are grouped together and an ablation study is conducted using U-Net models.

PCA is designed to reduce the number of bands by transforming them into smaller sets of uncorrelated variables, known as Principal Components. Each PC is a linear combination of the bands that captures as much of the explained variance as possible (Abdu, 2019). PCA assists in feature selection and help reduce the model's dimensionality by suggesting a number of principal components that explain 95% explainability of the variance. In the case of pixel-level classification or segmentation, such as land use classification, PCA will not be used to create its own features; instead, it assists in feature selection by providing an indication of the number of components that could be used as inputs during modeling. PCA is conducted using the PCA package from

scikit learn in Python with all of the features created in Section 3.4.1-3.4.3, such as QA, regional, indices, and GLCM from Landsat imagery. The output of PCA is a cumulative explained variance curve, which indicates how many components along the curve explain what percentage of the variance. This can, in turn, be used as an indication of the number of features that could be used for the model-building process.

The second step is model-based explainability using SHAP. SHAP uses simple models as an interpretable approximation of the explainability of deep learning models by using an additive feature attribution method (Lundberg & Lee, 2017). The surrogate model g is defined as:

$$g(z') = \varphi_0 + \sum_{i=1}^{M} \varphi_i z'_{i'} \tag{15}$$

where $z' \in \{0, 1\}M$, M is the number of simplified input features, and

$$\varphi i \in R$$

in the case of this praxis, the simple model used is XGBoost. SHAP provides a per-feature importance score that is applied at the overall and per-class level, which are used to remove features with a small model explainability value.

The third step is Gini ensemble importance using random forest. Random forest provides a ranking of features by the average reduction in node impurity (Gini) for each predictor in the model. Gini can overestimate the importance of variables with highly dimensional and highly correlated data (Genuer et al., 2010). To alleviate this issue, a correlation matrix is used in the fourth step in iteration with the Gini and SHAP values. The addition of the correlation matrix helps by removing highly important, highly

correlated features from the training data. Not only could highly correlated features overestimate importance, but they could also confuse models, degrading their accuracy.

Lastly, the ablation study is conducted to ensure feature selection. While feature importance was conducted using RF, ablation studies were conducted using U-Net and Vision Transformers (ViT) to ensure RF and XGBoost model feature importance are applicable between the differing architectures. Features are grouped by two criteria: (1) acceptance or rejection of the features and (2) their primary source. For example, this would result in (1) accepted indices and (2) rejected indices. Drop-one ablation is conducted first, using all groups of data, dropping one at a time and evaluating the impact. Next, only one ablation is conducted, running the model on only that group to identify the impact. The ablation study is the last step in the feature engineering process, confirming all features that support model accuracy in accordance with the model evaluation metrics in §3.6

3.5 Model Training and Optimization

Model training is needed to test the effectiveness of the selected state-of-the-art models, and optimization is completed to maximize the accuracy of each of those models. These two functions are designed to test RQ1-RQ4 and H1-H4, as this praxis is testing (1) with each model maximized for accuracy, how they compare, and (2) whether temporal features aid in that accuracy. This analysis helps to understand the current state of the art machine learning models for land use classification.

All models were trained using the output described in 3.4.4. The data is prepared for use by all models, with separate train, test, and validation file sets. Each model simply

loads this data for identical inputs and outputs. All model training, epochs (where applicable), and test evaluations are standardized and recorded for each model and outlined in §3.6.

Decision tree-based models are used to set baseline accuracy in line with the most recent NLCD accuracy assessment (Wickham et al., 2023). The U-Net model is currently used by the NLCD, even though no published accuracy assessment. Due to the recent adoption of U-Net, this praxis explores its effectiveness against RF models in deep networks for forest accuracy gains (RQ1 and H1) (Brown, 2024, p. 2). With expanding popularity among remote sensing use cases, ViT have been outperforming U-Net and other convolutional neural network (CNN) backed models (Yao et al., 2023). Data-efficient image transformer (DeiT) is a ViT that is chosen for this praxis due to its lightweight architecture (Han et al., 2023). A comparison of U-Net models and ViT addresses vision transformers for accuracy gains (RQ2 and H2). Hybrid CNN + Transformer models have been extensively tested on medical image segmentation with great success (Y. Zhang et al., 2021). Hybrid models have yet to have the same extensive testing or usage in remote sensing because of their more recent development. This praxis tests a STransU2Net structure, which has been shown to outperform U-Net models by up to 5% (G. Liu et al., 2024). This hybrid model can be compared against U-Net to address hybrid transformers for accuracy gains (RQ3 and H3).

Model training is conducted using Lambda Labs, a cloud provider of graphical processing units (GPUs) servers. Lambda Labs offers instances with state-of-the-art GPUs. These instances offer up to 208 virtual CPUs and 1800 gigabytes of RAM. Lambda Labs offers a secure cloud connection allowing code to be executed from

Microsoft Visual Studio Code and cloud filesystems that can be connected to the H100 GPU instances for fast training. For deep learning architecture TensorFlow and Keras application programming interface (API) libraries are used within Jupyter notebooks.

3.5.1 Tree-Based Models

Tree-based algorithms are used as the baseline models for this praxis. In land use classification tree-based models are particularly well-suited for high dimensional pixel level multi class classification. Both the random forest (RF) and extreme gradient boosting (XGBoost) models are trained on pixel-level values from the final feature set using the same training, validation, and test dataset as the rest of the models outlined in §3.5. Both the RF and XGBoost models support deep networks for accuracy gains (RQ1 and H1) as the baseline model to be used in comparison with U-Net models. Setting a baseline model RF and XGBoost help validate that the data used to train the models can in fact be useful for lulc modeling. RF was specifically selected because it is the last model used by the NLCD with a published accuracy assessment (Wickham et al., 2023). Tree-based models have been used by the NLCD from 2001 to 2019; for this reason, a second tree-based model was selected to confirm baseline accuracy. XGBoost was chosen as this second model because it is a well-supported and documented tree-based model with GPU support for fast training.

Historic NLCD publications show usage of decision trees and in 2019 random forest, the details surrounding the architectural choices and hyperparameters and not discussed. The RF model used is loaded from the sci-kit learn package. For the purpose of this praxis the hyperparameter search includes maximizing accuracy over n_estimators, max_depth, min_samples_leaf, min_samples_split, max_features, and

class_weights. The other models used in this praxis are dense segmentation models which use the entire image to detect class distribution and spatial co-occurrence. To ensure comparability of the segmentation models with the RF baseline this praxis does not re-balance pixels using oversampling or undersampling techniques. Instead imbalances are handled by including class weights provided to the model using scikit learns utility compute_class_weight based on the formula below. For class c, the weight is where N is

$$\boldsymbol{w_C} = \frac{\boldsymbol{N}}{\boldsymbol{K}\,\boldsymbol{N_c}} \tag{16}$$

the total number of samples, K is the number of classes, and N_c is the number of samples in class c. XGBoost follows the same class imbalance handling as RF models, but the class imbalance is handled using sample_weight which is the XGBoost equivalent of compute_class_weight. A full list of parameter and hyperparameters tested and the optimally selected values are outlined in §4.5.1.

3.5.2 U-Net

U-Net was selected as a state-of-the-art model as it is the currently used in production by the MRLC to produce the NLCD. However, no published accuracy assessment of this new architecture. A U-Net model addresses the model architecture comparisons (RQ1-3 and H1-3) (see §1.6) as the architectures are designed for semantic segmentation and work to capture the global context and skip connections that maintain fine spatial details that exist within land use classification. What makes U-Net models unique is the symmetrical encode-decoder network with pixel-wise labeling. The current NLCD implementation of U-Net classification algorithm consists of 256x256 multi-band tiles and is down-sampled to 16x16, which influenced the window size of the sampling

method. Spectral indices are also included such as NDVI and NDWI. The current methodology of the NLCD includes a much more complex architectural design with multiple models at different stages focused on specific aspects of land cover such as impervious surfaces or the national wetland index. Two spatial U-Nets are used to assist in land cover classification. One model focused solely on land cover uses a Focal Jaccard loss function and was built on 9.8 million parameters. The second model assists with urban intensity, covering many urban classes and uses a Focal Dice loss function and is built on 26.2 million parameters (Brown, 2024, p. 2). The focal loss mitigates class-imbalance by down-weighting easy, high-confidence examples and focusing on low-confidence or misclassified samples. Dice maximizes the overlap between the prediction and ground truth, raising recall for minority classes. (Abraham & Khan, 2019). The Jaccard index is an intersection over union (IoU) score, which gives the ratio of predicted values and ground truth at their union (Berman et al., 2018). Other Training applications noted model convergence up to 600 epochs with 35 steps per epoch and learning rates are noted from 0.000005 to 0.00003 (Freudenberg et al., 2019).

To align with the NLCDs implementation the U-Net for this praxis is designed to process leaf-on and leaf-off layers in parallel. This two-stream method uses all surface reflectance bands, thermal, DEM, spectral indices, and textural features with their associated leaf-on or leaf-off layers. Training loss function candidates include weighted cross-entropy, focal variants, Dice, Jaccard/Iou, and combinations of each. A full list of tested architectural characteristics, parameters, and hyperparameters, along with the selected values, is documented in §4.5.2.

3.5.3 Vision Transformer

Vision Transformers (ViTs) natively capture long-range context using self-attention rather than the local convolutions, pooling, and dropouts from U-Net. ViTs directly address vision transformers for accuracy gains (RQ2 and H2) as a comparison model against U-Net. Unlike its U-Net / Convolutional partner ViTs use self-attention mechanisms which can model global dependencies without spatial bias that may interfere specifically with continuous level 2 classification such as forest type or urban density. To reduce the overall computational complexities of ViT's this praxis explores the use of light ViT variant models such as Data-efficient Image Transformers (DeiT) or shifted window (Swin) Transformers. Keeping the models minimal computationally is in line with the NLCD priority of maintaining a scalable nature for its modeling.

This praxis explores various pretrained transformer variants such as DeiT, Swin, and SegFormer to assess their applicability to the NLCD data. Most published documented use of these models is often accomplished on smaller geographic zones, fewer target classes, and with cleaner input data. The models are applied for dense segmentation by dividing 64x64 tiles into 8x8 patches. Transformer blocks provide global multi-head self-attention (MSA) and a multi-layer perceptron (MLP) feed-forward step. Predictions are up-sampled back to a 64x64 grid to produce per-pixel labels. Similar to U-Net class weighted cross-entropy, focal variants, Dice, Jaccard/Iou, and combinations of each are explored with the full list of tested architectural characteristics, parameters, and hyperparameters along with the selected values are documented in §4.5.3.

3.5.4 Hybrid CNN + Transformer

Hybrid CNN + Transformer models combine local featurization of the CNNs with the global attention of transformers. These architectures, when used together, are particularly well-suited for the complex task of remote sensing classification. This praxis utilizes a STransU2Net architecture to address research question 3 and hypothesis 3 (see §1.6), integrating transformer blocks within the U-Net framework. By combining these architectures, the model can be better equipped to resolve spatial ambiguities with historically low accuracy, such as class boundary lines between wetlands and forests or urban areas and agricultural areas.

STransU2Net integrates transformer blocks on the encoding side of the U-Net while decoder uses the same multi-scale up sampling with skip connections used in the traditional U-Net described in §3.5.2. Transformer tokens are created by patching the encoder feature maps and learning relationships among the patches. These fused features are passed back to the U-Net decoder for pixel-level dense segmentation. While the models were previously created some hyperparameter tuning are utilized to optimize the model. Similar to U-Net and ViT focal loss variants, Dice, Jaccard, and combinations of each are explored with the full list of tested architectural characteristics, parameters, and hyperparameters along with the selected values are documented in §4.5.4.

3.5.5 Hyperparameter Tuning

This praxis used Optuna to perform hyperparameter tuning. It was used on U-net, Vision Transformer, and Hybrid CNN + Transformer architectures. Specific models are deployed with Optuna to automate the hyperparameter search and optimization to locate optimal hyperparameters efficiently. Optuna uses an adaptive Tree-Structure Parzen

Estimator (TPE) sampling strategy based on Bayesian optimization rather than traditional grid search which can be computationally time consuming, especially with the more deep neural networks (Akiba et al., 2019) .

3.6 Model Evaluation

A standardized and model-agnostic evaluation technique support consistent evaluation of the model outputs. All of the final metrics are delivered at the Level 1 classification, Level 2 classification, and overall level: Accuracy, Balanced Accuracy, and Mean intersection over union (mIoU).

3.6.1 Multi-Class Classification Evaluation Metrics

All four of the research questions in this praxis are related to model accuracy and comparison. LULC applications require multi-class classification, typically using a dataset with unbalanced classes. The purpose of LULC classification is to match spectral bands to information classes (J. Campbell & Wynne, 2011). The NLCD uses 8 level 1 classes and 16 level 2 classes (Brown, 2024).

The NLCD uses overall accuracy as their core metric where $\widehat{\boldsymbol{P}}_{\boldsymbol{h}}$ is the share of correctly classified pixels in stratum h, N is the total number of sample pixels, $\boldsymbol{N}_{\boldsymbol{h}}$ is the population size of stratum h, and the summation is over all H strata (H = 22) (Wickham et al., 2023). This finds the probability that a random land use class is correctly weighted by its true share of the overall population. For comparison against the NLCD this metric must be used.

$$\widehat{\boldsymbol{OA}} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{h}=\mathbf{1}}^{\boldsymbol{H}} \boldsymbol{N}_{\boldsymbol{h}} \widehat{\boldsymbol{P}}_{\boldsymbol{h}} \tag{17}$$

Overall accuracy can mask poor performance for rare classes since it is weighted with respect to its total population. Using this measure directly would highlight specific challenges with the NLCD such as the 50% accuracy assessment for the minority snow classification (class 12). For model training, this praxis uses three additional metrics to better capture overall performance: Accuracy, balanced accuracy, and Mean intersection over union (mIoU).

Accuracy shows the percentage of how many of the predicted values are actual values, giving a declared probability that the model is correct (Grandini et al., 2020). Accuracy is one of the most widely utilized evaluation metrics when experiencing a balanced multiclass classification problem (Chicco & Jurman, 2020) as:

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP + TN}}{\boldsymbol{TP + TN + FP + FN}} \tag{18}$$

where true positive (TP) is the number of samples correctly predicted as belonging to a class, true negative (TN) the number of samples correctly predicted as not belonging to a class, false positive (FP) the number of samples incorrectly predicted to belong to a class, and false negative (FN) the number of samples that belong to a class but were incorrectly predicted as another class.

When choosing the highest performing model, balanced accuracy is taken into consideration to account for low-representation classes. Balanced accuracy gives importance to each class based on its frequency, giving them a proportional effect in the accuracy score (Grandini et al., 2020). To measure imbalanced data with multi-class classification, balanced accuracy needs to incorporate macro precision and macro recall

by averaging per-class cores, together creating a balanced class-wise accuracy score (Beddar-Wiesing et al., 2025).

$$\boldsymbol{TNR_{macro}} = \frac{\mathbf{1}}{\boldsymbol{C}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{C}}\frac{\boldsymbol{TN_i}}{\boldsymbol{TN_i + FP_i}} \tag{19}$$

$$\boldsymbol{TPR_{macro}} = \frac{\mathbf{1}}{\boldsymbol{C}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{C}}\frac{\boldsymbol{TP_i}}{\boldsymbol{TP_i + FN_i}}$$

$$\boldsymbol{Balanced\ Accuracy} = \frac{\boldsymbol{TPR_{macro} + TNR_{macro}}}{\mathbf{2}}$$

Mean intersection over union (mIoU) range for 0-1 indicating overlap between predicted and ground truth labels, higher mIoU values indicate better semantic segmentation performance. It is also known as Jaccard Index and it provides a well-accepted measure of how well a segmentation model detects the boundaries and regions of land classes (Dekker et al., 2023). While intersection over union (IoU) is a common metric for segmentation tasks, multilabel classification is more commonly averaged, resulting in the mean IoU or (mIoU) (Berman et al., 2018).

$$\boldsymbol{Mean\ IoU} = \frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\boldsymbol{IoU\ class\ i} \tag{20}$$

Chapter 4 – Results

4.1 Introduction

This praxis is centered around the development of a state-of-the-art model for land use and land cover (LULC) classification with an accuracy that outperforms the National Land Cover Database's (NLCD) 77.5% benchmark (Wickham et al., 2023) by achieving at least 90% accuracy. The results from this chapter follow the methodology outlined in the preceding chapter. Deep networks for accuracy gains (RQ1 and H1) compare U-Nets to random forest models in differentiating forest types for the continental United States (CONUS). Vision transformers for accuracy gains (RQ2 and H2) compare vision transformers (ViT) to U-Net models for overall LULC classification. Hybrid transformers for accuracy gains (RQ3 and H3) compare hybrid CNN + transformer models in comparison to U-Net models in overall LULC classification. Finally, temporal context for accuracy gains (RQ4 and H4) compares temporal mechanisms for accuracy, specifically for level 2 areas.

Section 4.2 describes the acquired datasets for the Landsat scenes, digital elevation model (DEM), NLCD land cover and confidence, and the Moderate Resolution Imaging Spectroradiometer (MODIS) phenological data. Section 4.3 provides descriptive statistics and graphical diagnostics of the data used to influence the feature engineering steps. Section 4.4 details the engineering of new features, as well as the analysis of the importance of indices, textural, and temporal data. Model evaluation in §4.5 compares the performance of the candidate models described in §3.5 using the standardized metrics outlined in §3.6.

4.2 Data Acquisition

The data selected for this praxis is designed to mirror the current practices of the multi-resolution land characteristics (MRLC) while enhancing it, where possible, for increased reusability. In line with the expected output outlined in §3.2 collectively, the sources listed in Table 4.2a below resulted in 785 gigabytes (GB) of data, consisting of 3,706 2048x2048 40-band tagged image file format (TIFF) files.

Table 4.2a - Data Acquisition

Dataset	Source	Resolution	Key Layers Included
Landsat 8	USGS	30m	Surface Reflectance, QA, and Surface Temperature
NLCD Land Cover	USGS	30m	Land Cover Classification (Labels)
NLCD Confidence	USGS	30m	Land Cover Confidence Scores
NASA DEM	NASA	30m	Elevation, Slope, Aspect, and Hill Shade
MODIS	NASA	500m	Phenological Dates
EPA Ecoregion	EPA	N/A	Ecoregion Zone 1, Zone 2

The NLCD Land Cover is organized hierarchically into 8 level 1 classes and 16 level 2 classes. The land cover classes and pixel population from the raw input data, excluding invalid pixels, are outlined in Table 4.2b below. Within the NLCD land cover classification, four classes do not appear in the CONUS-focused praxis dataset because the classes only appear in Alaska: (1) Class 51 Dwarf Scrub, (2) Class 72 Sedge/Herbaceous, (3) Class 73 Lichens, and (4) Class 74 Moss.

Table 4.2b - Land Use Classification

Level 1	Level 2	Land Cover Class	# of Pixels	% of Total Pixels
10	11	Open Water	489,674,112	5.32%
10	12	Perennial Ice/ Snow	995,891	0.01%
20	21	Developed, Open Space	247,695,578	2.69%
20	22	Developed, Low Intensity	158,020,948	1.72%
20	23	Developed, Medium Intensity	95,109,901	1.03%
20	24	Developed, High Intensity	34,133,483	0.37%
30	31	Barren Land	84,907,065	0.92%
40	41	Deciduous Forest	846,722,232	9.40%
40	42	Evergreen Forest	1,068,313,352	11.62%
40	43	Mixed Forest	316,342,899	3.44%
50	52	Shrub	1,959,136,399	21.30%
70	71	Grassland/Herbaceous	1,226,584,167	11.34%
80	81	Pasture/Hay	581,516,105	6.32%
80	82	Cultivated Crops	1,509,342,850	16.41%
90	90	Woody Wetlands	418,425,617	4.55%
90	95	Emergent Herbaceous Wetlands	142,293,229	1.55%

4.3 Data Preprocessing and Exploratory Analysis

The data preprocessing and exploratory analysis are conducted on the merged dataset output from the processes in §3.2, which contains 40 bands of 3,706 2048x2048 pixel images. The goal of the exploratory data analysis (EDA) is to understand the characteristics of the dataset, while the data preprocessing steps segment and optimize the data for feature engineering and selection.

Two quality bands that are crucial to the data in this praxis. First, the confidence score gives an indication of how confident the current land classification is by the NLCDs model, acting as a quality score in the ground truth label. As seen in Figure 4.3a, classes such as water (11), high intensity development (24), cultivated crops (82), and woody wetlands (90) have relatively high average confidence scores. On the other hand, snow (12), medium-intensity development (23), low-intensity development (22), open space development (21), and mixed forest (43) have lower confidence scores, suggesting possible issues in labeling accuracy. Of highest concern is the open space development (21) because 75% of the pixels are under 60% confidence, and it is a minority class.

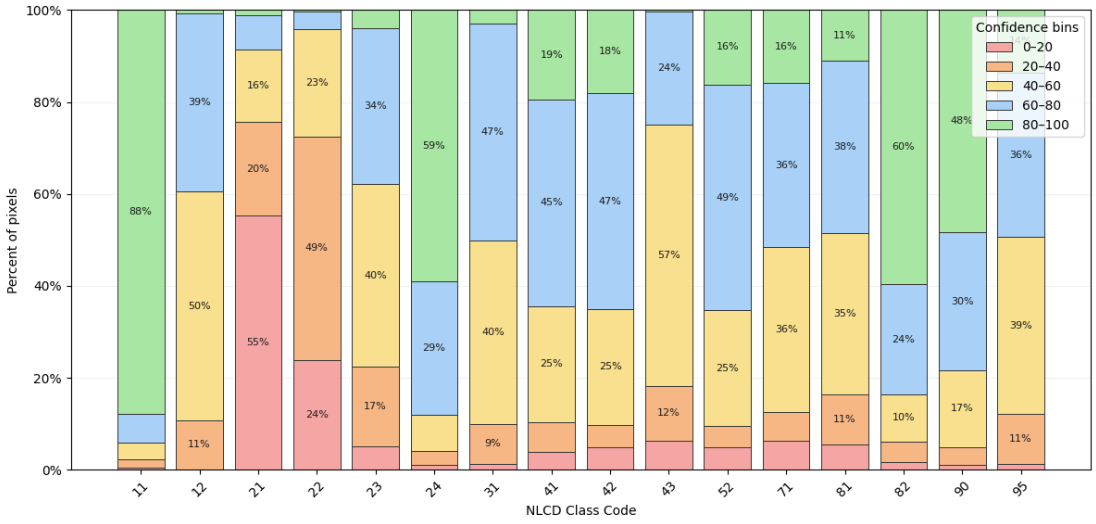

Figure 4.3a - Confidence Bins

The second quality band of high importance is the quality assurance pixel band (QA_PIXEL) from Landsat, as seen in Figure 4.b. For each class, a leaf on (left) and leaf off (right) bar displays the percentage of QA categories that make up that class. All the classes except water (11) and snow (12), clear pixels are ideal, and as data is filtered, this should improve. For class water (11), either water or clear is an acceptable QA class.

Regarding class snow (12), either water, snow/ice, or clear would be an acceptable QA class. Although all classes contain cloud, shadows, and fill, class snow (12) is the most affected by this. Paired with its lower-than-average confidence scores and low relative pixel count, this is a class of concern while building the model dataset.

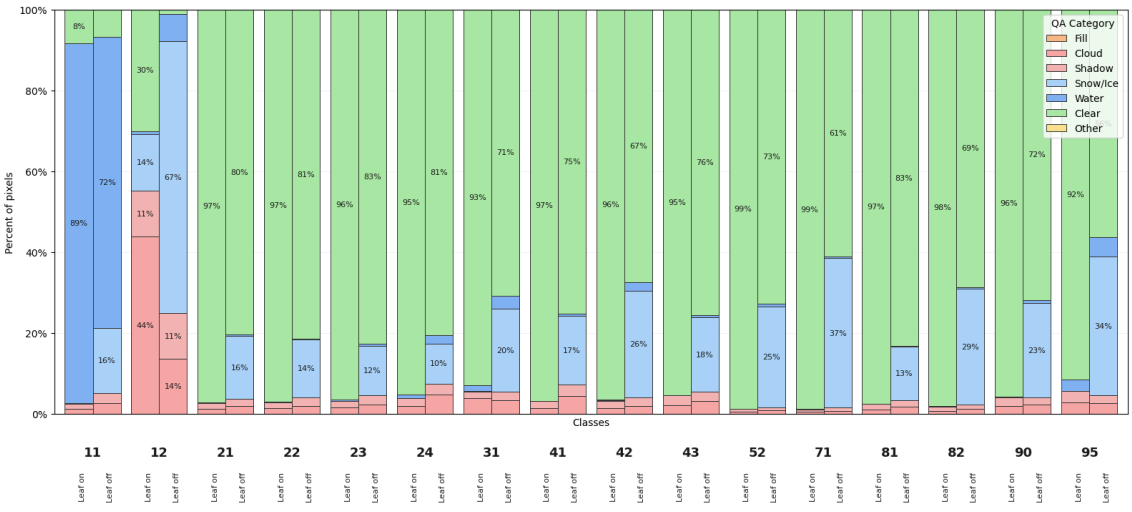

Figure 4.3b - QA Pixel Category

4.3.1 Adaptive Window Method

The adaptive window method was designed to accommodate the differences in pixel counts, confidence scores, and Landsat quality. Outlined in §3.3.1, this sampling method allows for increasingly lower confidence thresholds and oversampling pixel areas to increase the rare class population. While running the algorithm detailed in §3.3.1 for each 64 pixel image, the stride, confidence threshold, pixel count and the percentage of pixels retained from the raw data are outlined in Table 4.3.1 below. Percent of total pixels includes but does not list invalid pixels, percent of pixels retained can be higher than 100% due to the resampling method. The sampling method proved useful in increasing the total percentage of pixels for classes 12, 31, 90, and 95. The developed land classes,

on the other hand, were penalized due to the lower percentage of confidence scores for those classes. During the adaptive window method, summary statistics for minimum, maximum, mean, standard deviation, first quartile, and 3rd quartile values were recorded for each band.

Table 4.3.1 - Adaptive Window Results

Class	Stride	Confidence Threshold	Pixel Count	Percent of Total	% of Pixels Retained
11	64	80	172,629,907	8.38	35.25
12	4	60	38,594,974	1.87	3,875.42
21	64	80	22,784,823	1.11	9.20
22	64	80	13,099,749	0.64	8.29
23	64	80	7,585,345	0.37	7.98
24	64	80	5,242,194	0.25	15.36
31	32	75	210,328,793	10.21	247.72
41	64	80	113,696,938	5.52	13.43
42	32	75	316,822,504	15.38	29.66
43	32	75	12,680,812	0.62	4.01
52	64	80	282,704,704	13.72	14.43
71	64	80	104,784,684	5.09	8.54
81	32	75	54,722,555	2.66	9.41
82	64	80	472,967,813	22.95	31.34
90	64	80	170,753,583	8.29	40.81
95	64	80	61,122,094	2.97	42.96

While a reduction of the majority classes (11,31,42,52,82) would be ideal, the interconnectedness of these classes with minority classes. Many of the classes appear in the same window as other classes, this mix of co-occurrence is seen in Figure 4.3.1a below. A total of 714,334 windows were removed to thin majority classes. The rest of

the TIFFS were kept to avoid missing training situations where these classes do appear individually.

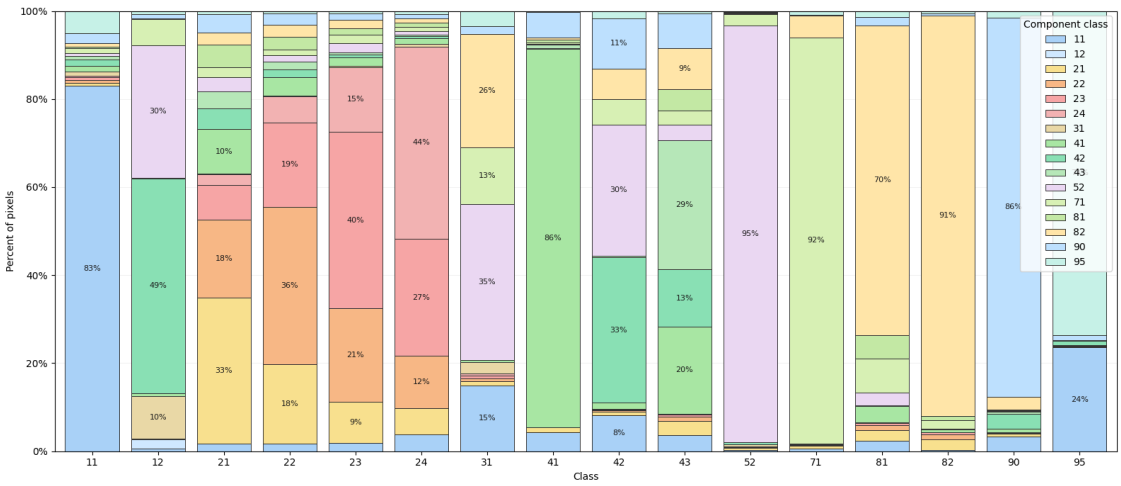

Figure 4.3.1a – Class Confusion

However, the purpose of the sampling method was not to balance the dataset, but to oversample rare classes and create a high-confidence dataset. The increase in average confidence scores can be observed in Figure 4.3.1b over the confidence scores in Figure 4.3a.

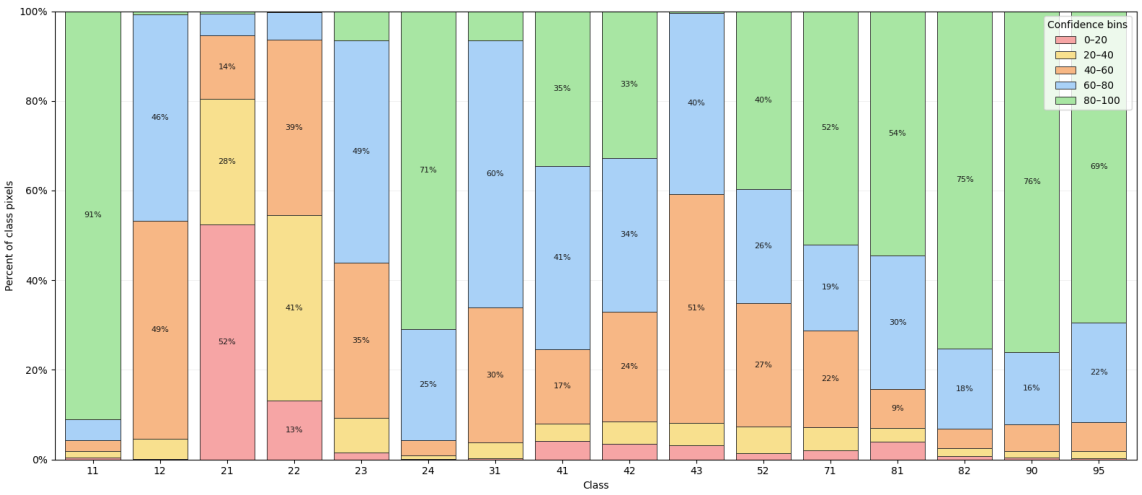

Figure 4.3.1b - Class wise Confidence Scores

4.3.2 Data Filtering

The total number of windows output from the adaptive window method is 1,320,764. Of these 7,918 contain N/A values for any pixel among any of the bands and were the first to be filtered out. Using QA_PIXEL pixel quality assurance values from Landsat removed 246,436 windows due to clouds (cloud, cirrus, dilated), snow pixels not on class 12 (snow), and cloud shadows. A confidence score filter removed 93 windows because their confidence score distribution had the first quartile below 40. While checking QA_RADSAT, radiometric saturation was checked and no windows were removed. The result of the quality assurance filters was reduction of used windows down to 1,074,235.

Of these 1 million windows 75% of them are represented by classes water (11), evergreen forest (42), shrub (52), and cultivated crops (82). A class specific filtering is conducted to both thin and increase confidence scores. For cultivated crops (82) 95% of windows containing only cultivated crops with less than 98% average confidence score

were removed. This resulted in the removal of 2,585 windows. The same logic was applied to shrub (52) that resulted in the removal of 20,219 windows, evergreen forest (42) removed 17,841 windows, and water (11) which removed 52,302 windows. Finally, 5,188 windows were removed.

4.3.3 Grouped Scene Data Split

Given the cleaned and optimized windowed TIFF files from the adaptive window sampling method, each area of interest (AOI) is then assigned to a train, validation, or test set. The target data split was 70/15/15, with the final training dataset containing 140 million pixels (70.5%), the validation dataset containing 30 million pixels (14.5%), and the test dataset containing 30 million pixels (17.3%). The spatial penalty and minority class targeting outlined in §3.3.3 appear to be effective, given the details in Figure 4.3.3 below. It is interesting to note the count of AOIs for the data splits as they do not fit the split criteria. Upon further investigation, this is due to the number of TIFFs per AOI

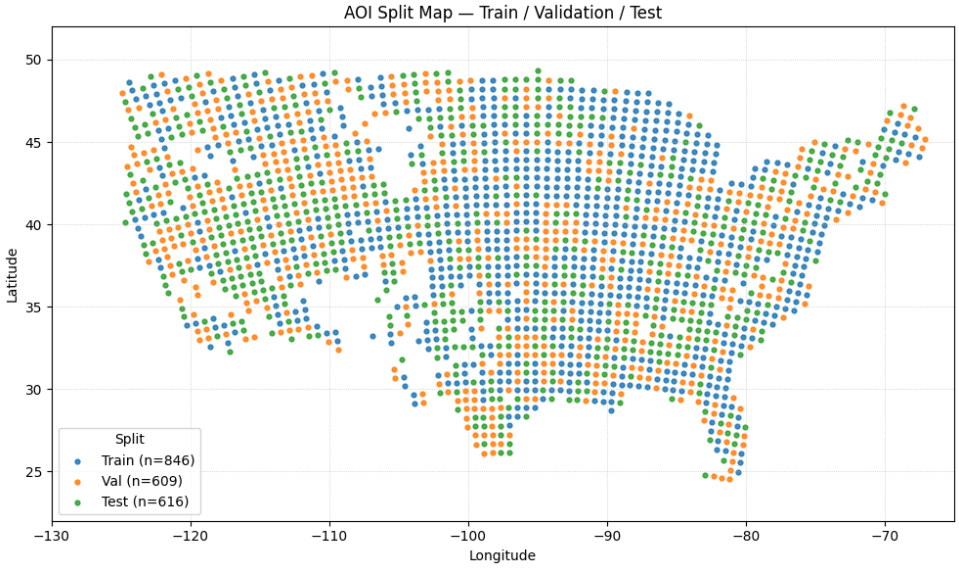

Figure 4.3.3 - Grouped Scene Data Split

being a higher density for training: 419 for training, 121 for validation, and 120 for the test set.

4.4 Feature Engineering

Following the feature engineering methodology outlined in §3.4 was designed to produce a data rich model agnostic input dataset for model training on random forest, U-Net, ViT, and hybrid models. First, 24 spectral and thermal bands are standardized to prevent scale-related errors, particularly in U-Net and ViT architectures. Next, 50 One-hot encoded quality assurance bands were created to drive model aware Landsat image characteristics. Then, 48 spectral indices were created to help drive model understanding of well-known mixed band classes such as those found in urban development. Additionally, 50 textural features were created to capture spatial patterns within the images to drive understanding of texture rich features such as those found in farmland. Finally, 34 ecological features were added to provide a regional context, allowing for distinguishability between taxonomically similar yet regionally distinct classes such as eastern and western deciduous forests.

4.4.1 One-Hot Encodings

In addition to indices and textural features, categorical information embedded within the dataset is transformed into usable model inputs. This approach converts both text string and bit-packed information into explicit binary features to test for potentially important model features. The EPA ecological regions were created to handle regional domain shift (C9). Twenty-one level-2 EPA ecological regions were used, each large enough to prevent data leakage while still capturing regional variability. On average the regions contain approximately 104 AOI's and roughly 190 million pixels.

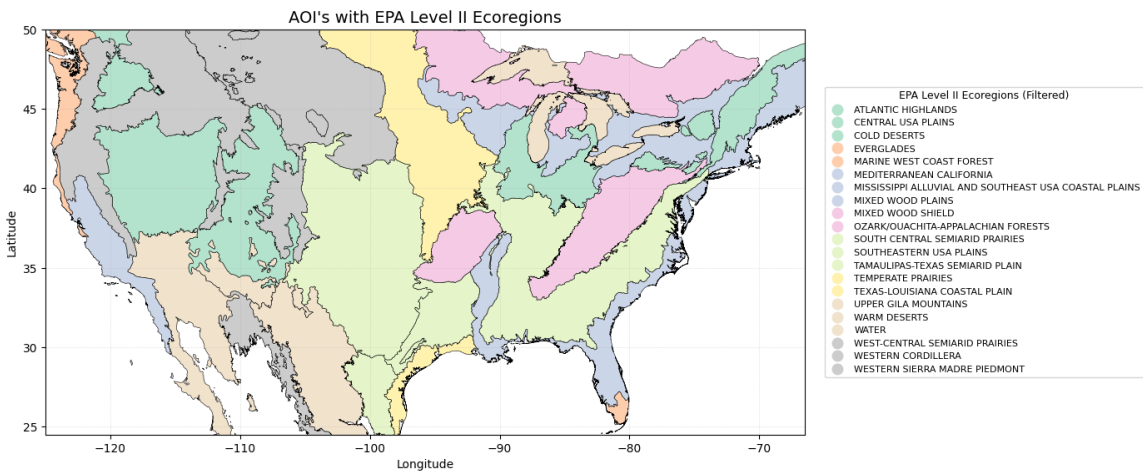

Figure 4.4.1 - EPA Ecoregions

Quality assurance encodings were created to take advantage of Landsat sensor readings. The analysis ready data provided by Landsat includes QA_Pixel, QA_Aerosol, and QA_Radsat bit-packed layers. Bands in Table 4.4.1 below are duplicated for both leaf-on and leaf-off layers.

Table 4.4.1 - Quality Assurance One Hot Encoded Layers

Source Band	New Band Name	Description
QA_PIXEL	QACAT_FILL	Pixel is image data or fill
QA_PIXEL	QACAT_CLOUD	Contains dilated, cirrus, or cloud values 0-1
QA_PIXEL	QACAT_SHADOW	Cloud-shadow confidence
QA_PIXEL	QACAT_SNOWICE	Snow/ice confidence
QA_PIXEL	QACAT_WATER	Water indicator
QA_PIXEL	QACAT_CLEAR	Clear condition
QA_PIXEL	QACAT_OTHER	Not clear condition or any other QA category
QA_RADSAT	RADSAT_ANY_SAT	Radiometric saturation status in any band
QA_RADSAT	RADSAT_B1_SAT	Radiometric saturation status for band 1
QA_RADSAT	RADSAT_B2_SAT	Radiometric saturation status for band 2
QA_RADSAT	RADSAT_B3_SAT	Radiometric saturation status for band 3

QA_RADSAT	RADSAT_B4_SAT	Radiometric saturation status for band 4
QA_RADSAT	RADSAT_B5_SAT	Radiometric saturation status for band 5
QA_RADSAT	RADSAT_B6_SAT	Radiometric saturation status for band 6
QA_RADSAT	RADSAT_B7_SAT	Radiometric saturation status for band 7
QA_RADSAT	RADSAT_B9_SAT	Radiometric saturation status for band 9
QA_RADSAT	RADSAT_TERRAIN_OCCLUSION	Line-of-sight blocked by terrain
QA_AERO	AERO_FILL	Pixel aerosol fill indicator
QA_AERO	AERO_VALID	Valid Aerosol Retrieval
QA_AERO	AERO_WATER	Water indicator
QA_AERO	AERO_INTERP	Aerosol value was interpolated
QA_AERO	AERO_LEVEL_CLIM	Aerosol level climatology
QA_AERO	AERO_LEVEL_LOW	Aerosol level low
QA_AERO	AERO_LEVEL_MED	Aerosol level medium
QA_AERO	AERO_LEVEL_HIGH	Aerosol level high

4.4.2 Spectral Indices

Spectral indices are added to enhance the specific characteristics of land cover classes. The indices in Table 4.4.2 below are designed to increase model understanding of urban built up areas, vegetation, water, wetlands, and snow. Each of these indices were created for both leaf-on and leaf-off layers.

Table 4.4.2 - Spectral Index Features

Purpose	**Indices Created**
Urban built up areas to detect impervious surfaces and man made features.	Urban Index (UI), Normalized Difference Built-up Index (NDBI), New Built-up Index (NBI), Band Ratio for Built-up Area (BRBA), Normalized Built-up Index (NBI), Modified Built-up Index (MBI, and Built-up Area Extraction Index (BAEI).

Vegetation to capture chlorophyll and canopy structure.	Normalized Difference Vegetation Index (NDVI), Modified Soil-Adjusted Vegetation Index (MSAVI), Chlorophyll Index (CI green), and Enhanced Vegetation Index (EVI).
Spatial change vegetation	NDVI gradient change in X-direction, NDVI gradient change in Y-direction.
Water, wetlands, and soil moisture.	Normalized Difference Moisture Index (NDMI), Normalized Difference Water Index (NDWI), Modified Normalized Difference Water Index (MNDWI), Normalized Difference Snow Index (NDSI), ShadowMask (empirical water/shadow/snow thresholds).

NDVI is a common vegetation index. Figure 4.4.2 below depicts the joining of surface reflectance band 4 (red) and 5 (near infrared) into the index.

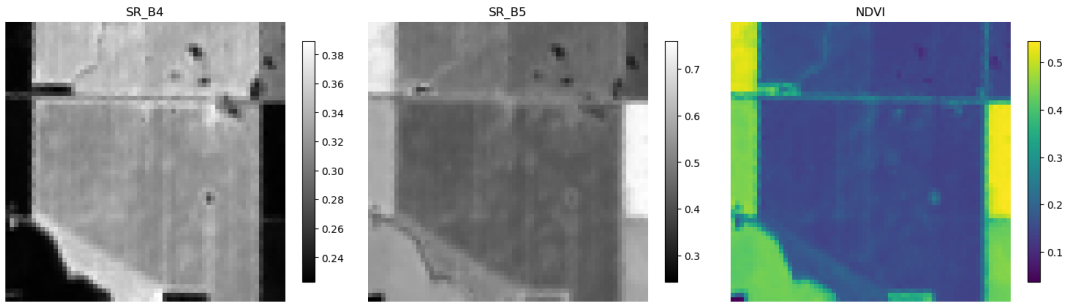

Figure 4.4.2 - Normalized Vegetation Index (NDVI)

4.4.3 Textural Features

Textural features are generated using the Gray-Level Co-Occurrence Matrix (GLCM) to capture spatial patterns not easily detectable in spectral features. For both

leaf-on and leaf-off seasons, five features were created: contrast, dissimilarity, homogeneity, energy, and correlation. As seen in Figure 4.4.3 below, these features highlight significant differences in the texture of a barren plateau near Colorado compared with a mountain desert in the north of Texas. Each of the five features were created for each of the five surface reflectance bands and for both leaf-on and leaf-off images resulting in an additional 50 bands.

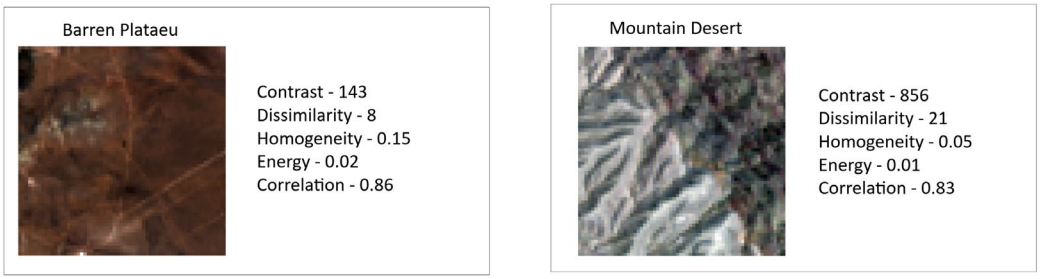

Figure 4.4.3 - Gray-Level Co-Occurrence Sample Values

4.4.4 Feature Selection

Feature selection was completed on 210 features across 16 million rows (1 million per class). As outlined in §3.4.4 PCA, Gini, Shapley Additive Explanations (SHAP), correlation matrix, and ablation studies are conducted to choose the minimum set of features that yield the most parsimonious model. The principal component analysis identified 90% cumulative explained variance (k) with 60 variables, 95% k with 75 variables, and 99% k with 107 variables. The elbow, or the point at which the learning curve for additional variables results in reduced k is 22 variables explaining 81% k. This praxis tested reducing the 210 features to between 75 and 107 features during ablation studies recorded later in this section. The PCA output informed the goal feature counts, capturing the majority variance for model parsimony while reducing dimensionality.

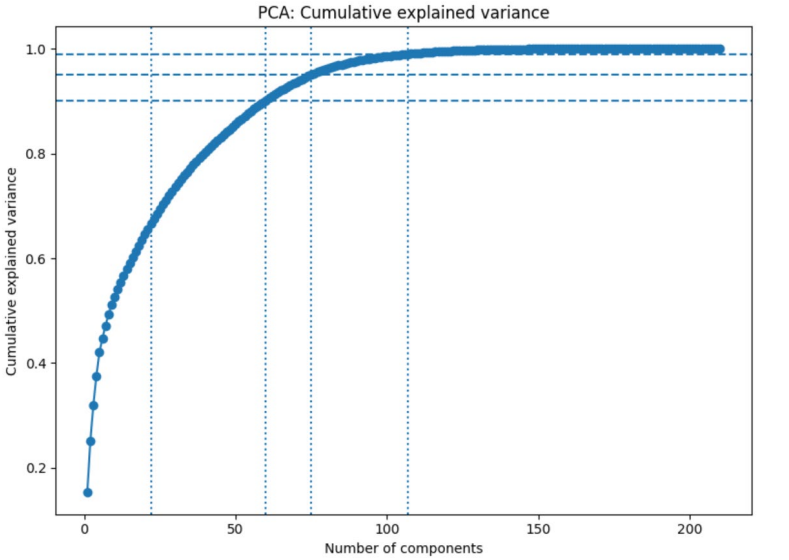

Figure 4.4.4a - Principal Component Analysis

SHAP was conducted both on the entire balanced dataset as well as binary for each class. Outside of the surface reflectance and temperature bands a few indices, one-hot encoded, and Gray-Level Co-Occurrence Matrix (GLCM) values had high SHAP importance. In Figure 4.4.2b below the high importance values for the overall feature set

are listed. In addition to this and in consideration of space, Table 4.4.2 summarizes the additional high importance values for binary per-class SHAP tests described in §3.4.5.

Table 4.4.4a - Binary Class SHAP Importance

Class	Band
11	TCB_STD
11	NASADEM_slope
11	MNDWI
11	NDVI_GRAD_Y
12	NASADEM_hillshade
21	NDVI_GRAD_X
21	UI
21	NDBI
22	MDWI_WINTER
52	EPA_L2_17
71	EPA_L2_6
82	EPA_L2_14

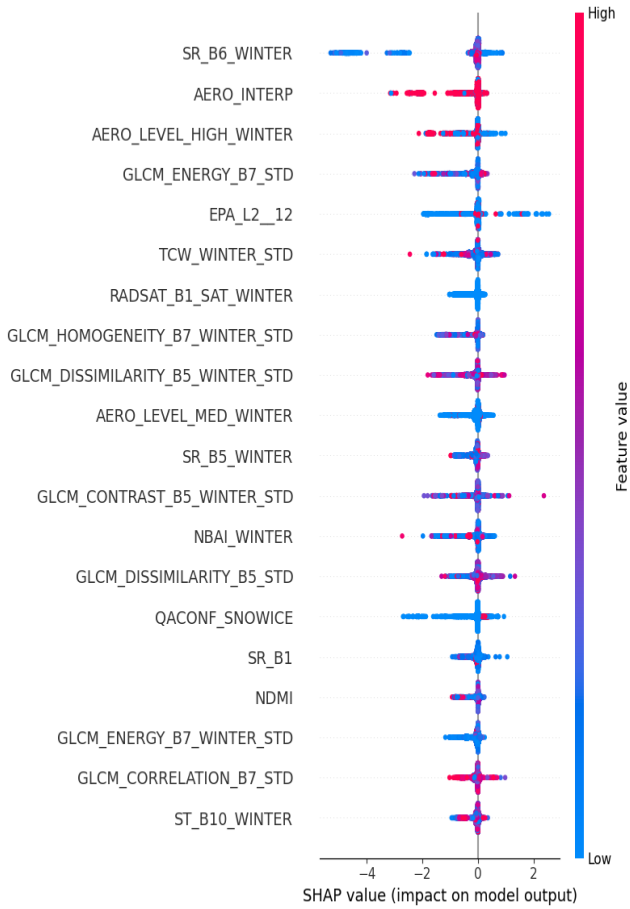

Figure 4.4.2b - SHAP Importance

Similar to SHAP, Gini importance was conducted on both the overall dataset as well as binary classification. NASA DEM elevation stood out as the most important feature by a significant amount. Figure 4.4.4c below identifies the vegetation-based indices, two impervious area indices, both leaf-on and leaf-off temperature, and tasseled

cap bands as some of the other high importance features. In addition to these features, additional features were added after consideration of the binary classification in Table 4

Table 4.4.4b - Binary Class Gini Importance

Class	Band
11	MBI
11	NDWI_WINTER
11	MSAVI
22	NBI
41	EPA_L2_10
52	EPA_L2_13
71	EPA_L2_19
71	EPA_L2_11
90	EPA_L2_9
95	EPA_L2_15

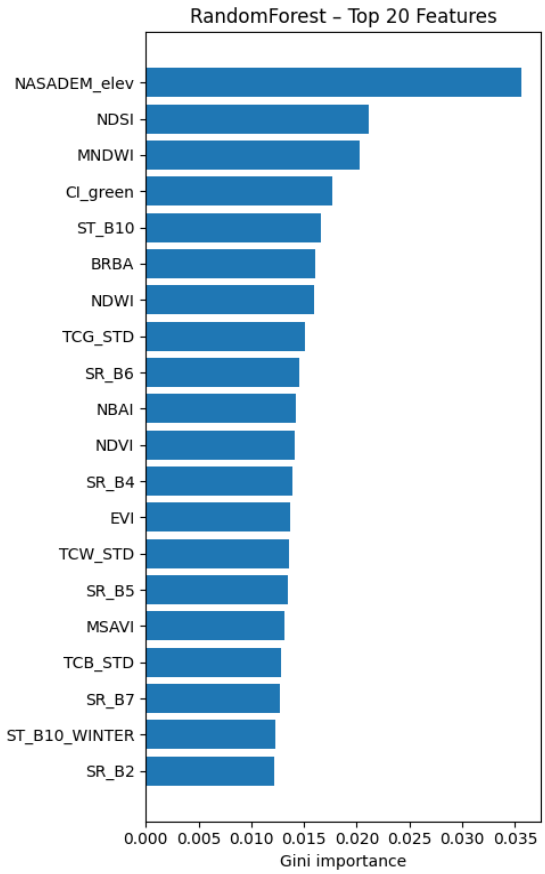

Figure 4.4.4c - Gini Importance

An analysis of the correlations between bands was conducted. This praxis identified that many of the indices had perfect or near-perfect correlation. The building indices NDMI, NDBI, and UI for both leaf-on/leaf-off were perfectly correlated. NDMI was kept out of the three due to its overall importance. MNDWI and NDSI were perfectly

correlated; NDSI was kept due to its overall importance. EVI and MSAVI had a 95% correlation; MSAVI was kept due to its higher significance in the Gini/SHAP analyses.

The ablation studies were conducted using U-Net and SegFormer ViT, which achieved a baseline accuracy of 86% and 90% respectively. The 210 features were grouped into 40 categories for computational efficiency such as NASA DEM, Surface Reflectance, Surface Temperature, Vegetation Indices, Building Indices, NDVI Graduals, Tasseled Cap, GLCM per band, Quality Assurance, Radiometric Saturation, Aerosol, and EPA Ecoregion. For each of these, there can be both an approved/unapproved version, as well as a leaf-on/leaf-off version. Initially, U-Net was the only ablation study conducted until the significant differences in the feature sets that drive accuracy were noted. A core characteristic of U-Net is its reliance on high-resolution features passed using skip connections for precise localization (Ronneberger et al., 2015). In contrast, SegFormer's hierarchical encoder and simple decoder fuse both high-resolution and low-resolution features (Xie et al., 2021). No feature set could be removed without a significant reduction in accuracy for at least one model. To ensure optimal model accuracy, different model types would have to be provided with features that align with the underlying architecture. The most prominent differences that occurred in feature sets that were originally recognized as important are outlined in Table 4.4.4 below.

Table 4.4.4 U-Net vs SegFormer Feature Importance

Feature Set	**U-Net Accuracy**	**SegFormer Accuracy**
Surface Reflectance - Removed	-1%	-42%
Approved Summer Indices – Only	73%	59%

Ecoregions – Removed	-10%	-28%
NASA DEM – Removed	-18%	-47%

4.5 Model

This praxis reports model results using a uniform pattern applied to the outputs for each model. All models are evaluated on identical train, validation, and test splits. First, baseline model performance is set in §4.5.1 with Random Forest and XGBoost models to simulate the 2019 NLCD performance. U-Net model results are recorded in §4.5.2 using the same model structures as the MRLC to estimate current performance in lieu of published accuracy assessments. Vision transformer results are outlined in §4.5.3 and hybrid CNN + transformer results in §4.5.4 as the state-of-the-art model tests.

4.5.1 Decision Tree-Based Models

Decision tree-based models provide this praxis with a baseline for all dense segmentation architectures to be compared against. Following the methodology in §3.5.1 a Random Forest (RF) and an Extreme Gradient Boosting (XGBoost) classifier were trained on the same data as the segmentation models. The RF and XGBoost hyperparameter search spaces and selected values are summarized in Table 4.5.1a and Table 4.5.1b below. To set a baseline, Random Forest needed to achieve an overall accuracy with 2019 NLCD weights in a relative range of 77.5%. Mean intersection over union (mIoU) should be used to compare RF model accuracy with that of the ViT and Hybrid models.

Table 4.5.1a Random Forest Hyperparameters

	Explored	**Selected**
Max depth	10-30	28
Number estimators	200-1000	829
Max features	Sqrt, log2	sqrt
Minimum samples leaf	1-6	1
Minimum samples split	2-12	8

Table 4.5.1b XGBoost Hyperparameters

	Explored	**Selected**
Learning rate	0.01-0.3	0.0541
Max depth	8-32	22
Minimum child weight	1-30	14
Column sampling per tree	0.6-1.0	0.67
Use class weights power	0.0-1.5	0.828

The 2019 NLCD accuracy assessment contained 3245 total test pixels. Nearly half (46%) of the pixels came from four of the sixteen classes, while the smallest class contributed just 0.12%; overall accuracy reached 77.5%, with average per-class accuracy of 83% for majority classes and 62% for minority classes. The complete per-class distribution and accuracy can be seen in Appendix Table A5. Using the same weights (H=22 strata) when computing Overall Accuracy (OA), the RF model achieves 78% accuracy, in line with expectations. As seen in Table 4.5.1c below, XGBoost attains higher OA (+6.5%) and mIoU (+0.513) due to its use of gradient-boosted decision trees. XGBoost has almost double the training times of RF, yet it is still under any of the deep neural network models.

Table 4.5.1c Tree-Based Model Outputs

Metric	Random Forest	XG Boost
Training Time	2721s	5381s
Mean Inference Time	5546ms	431ms
Overall Accuracy	78.6%	85.1%
Accuracy	80.1%	83.6%
Balanced Accuracy	77.7%	81.6%
mIoU	0.6727	0.7240

In testing for temporal context for accuracy gains (RQ-4 and H-4), RF models were tested with two feature groups. Temporal features improved both the accuracy and mIoU for random forest models relative to the non-temporal baseline, as seen in Table 4.5.1b below.

Table 4.5.1d Random Forest Temporal Results

	Overall Accuracy	Balanced Accuracy	mIoU
Including Temporal Features	78.6%	77.7%	0.6727
Not Including Temporal Features	69.1%	72.0%	0.5818

The confusion matrices show some misclassifications that occur at level 1 between Urban classes and Forest and Shrub vs Cultivated crops. Confusions between open urban and forest classes were noted in the 2019 NLCD, with a primary cause of remote area locations with unpaved roads being difficult to classify at 30 m resolution (Wickham et al., 2023). Level 2 classification accuracy shows some intra-level 1 class confusion. Most notably, this occurs in urban development between classes 21 and 22, then again in the forest between classes 43 and 41, as well as 41 and 42. An area of

concern is in especially low open-space developed and low-intensity developed land cover classes. The current MRLC process uses a model specifically for the urban development classification.

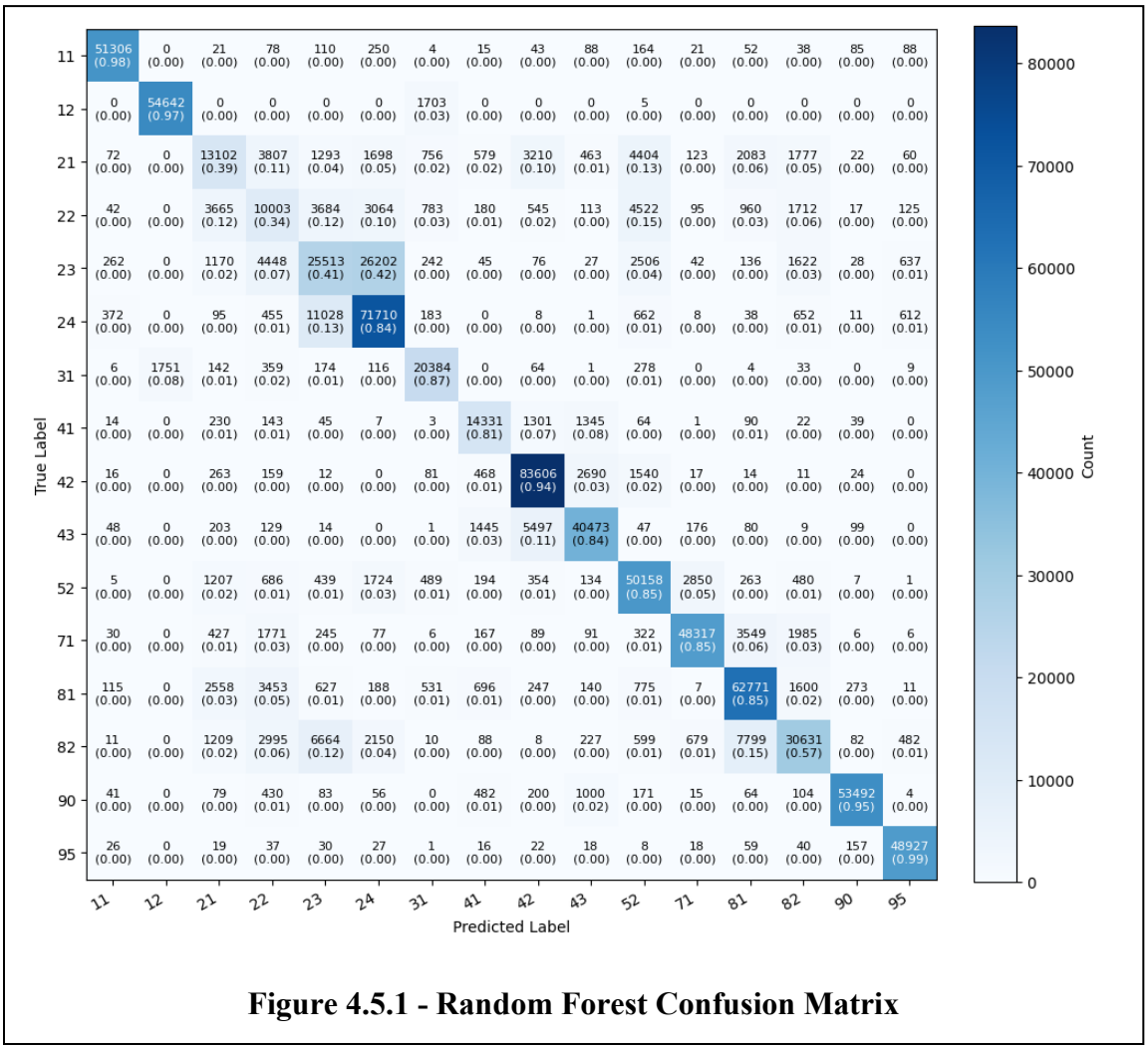

Figure 4.5.1 - Random Forest Confusion Matrix

4.5.2 U-Net

This praxis trained an encoder-decoder U-Net segmentation classifier in line with the current MRLC practices outlined in §3.5.2. This model is used in comparison with the RF model for deep networks for accuracy gains (RQ1 and H1) using overall accuracy and for vision transformers and hybrid transformers for accuracy gains (RQ2-RQ-3 and H2-

H3). The U-Net was trained on 100 epochs with a batch size of 16; the hyperparameters tuned are detailed in Table 4.5.2a below.

Table 4.5.2a U-Net Hyperparameter Search

	Explored	**Selected**
Optimizer	adam, adamw	adamw
Learning Rate	3e-5-3e-3	0.000125
Weight_decay	1e-6-3e-3	8.809
Base_width	48,64,80,96	96
Bottleneck_channels	512,768,1024,1280	512
Dropout_p	0.00-0.30	0.083
Loss_type	ce, ce+dice, dice, focal	Ce
Label_smoothing (ce loss type only)	0.00-0.15	0.148
Fuse_mode	Concat_1x1, sum	concat_1x1
Class_weight_power	0.0-1.5	0.874

The U-Net structure chosen for model development matches that of the current published NLCD process; however, this model requires dual seasonal input streams. To test the effect of temporal features, this praxis utilized a simpler single-input stream, which resulted in a higher mIoU as detailed in Table 4.5.2b below.

Table 4.5.2b U-Net Temporal Results

	Overall Accuracy	**Balanced Accuracy**	**mIoU**
Including Temporal Features	87.6%	83.5%	0.7346
Not Including Temporal Features	86.0%	84.3%	0.7478

While U-Net outperforms RF models in overall accuracy, it fails to support the deep networks for forest accuracy gains (RQ1 and H1). Between the three major forest classes (41, 42, 43), only a 0.6% increase in accuracy was observed, as seen in Figure 4.5.2a.

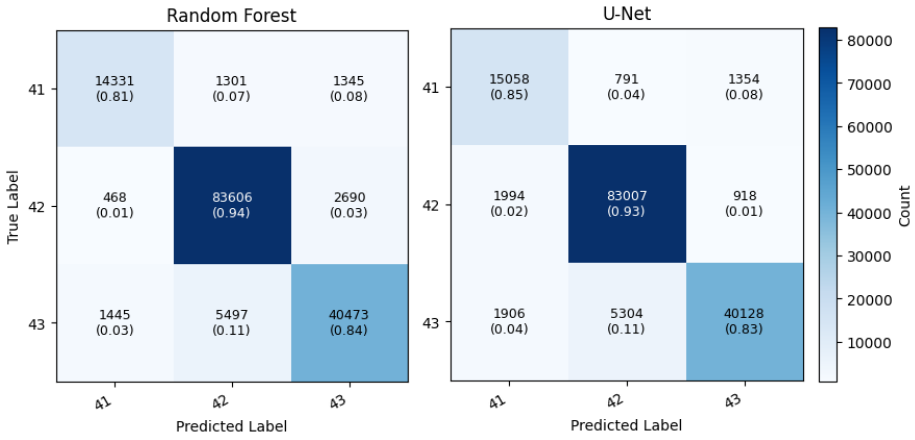

Figure 4.5.2a - U-Net Forest Class Confusion Matrix

The loss, accuracy, and mIoU curves in Figure 4.5.2b show a late convergence with approximately a 9% gap between train and validation. No signs of divergence as curves from all three metrics mirror each other, with validation being slightly rougher. The loss curve reaches a stable generalization by around epoch 95, confirming the fit for use of the architectural design for this dataset.

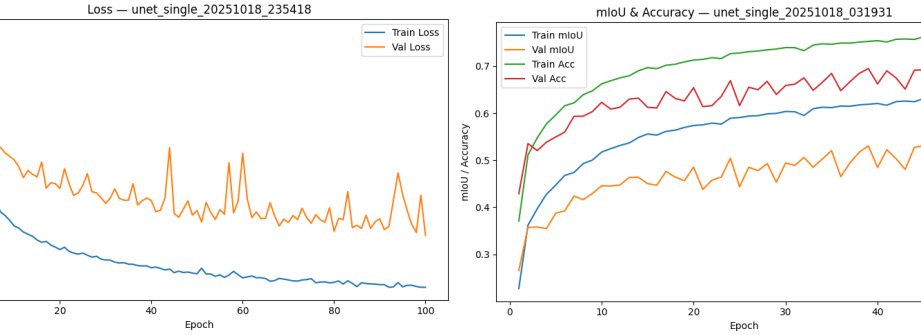

Figure 4.5.2b - U-Net Loss and Accuracy

4.5.3 Vision Transformer

SegFormer serves as the ViT's strongest dense segmentation classifier. Both Swin and SegFormer outperformed DeiT by 0.2 and 0.3 mIoU or 3-4% accuracy. Swin achieved the same accuracy as SegFormer, but SegFormer had a 0.05 increase in mIoU; therefore, it was selected. The SegFormer was trained on 100 epochs with a batch size of 96. The hyperparameter tuning searched the nine parameters outlined in Table 4.5.3a below and took 19 hours of computation to conduct. The difference between the best performing model and the worst was 0.19 mIoU or approximately 4% in accuracy. This significant change in mIoU was most notable in patch size, depth, and weight decay.

Table 4.5.3a SegFormer Hyperparameter Search

	Explored	**Selected**
Patch size	4,8,16	4
Embeded dimensions	192,256,320,384	256
Heads	4,6,8,12	8
Depth	4-10	4
MLP_mult	2-4	2.43
Weight_decay	0.01-0.2	0.1833
Dropout	0.00-0.30	0.1892
Attention Dropout	0.00-0.20	0.0547
Stochastic Depth	0.00-0.20	0.0308

Adding temporal features with SegFormer resulted in a 2.6% increase in overall accuracy, a 1.9% increase in balanced accuracy, and a .0226 increase in mIoU as detailed in Table 4.5.3b. ViT's global attention leverages the season context from leaf-on leaf-off pixels at a greater measurable increase over U-Net.

Table 4.5.3b SegFormer Temporal Results

	Overall Accuracy	**Balanced Accuracy**	**mIoU**
Including Temporal Features	88.5%	84.7%	0.7692
Not Including Temporal Features	85.9%	82.8%	0.7466

The confusion matrix in Figure 4.5.3a indicates a strong level 1 class precision. The lightest elements exist primarily in urban development (21, 22, 23), while high-intensity development remains higher at 86% accuracy. Although SegFormer has an overall increase in performance metrics, at a macro level it struggles with similar level 1 classes, albeit in a different way. SegFormer tends to struggle with more locally defined classes rather than the global classes highlighted in the U-Net experiments. This is discussed further in §5.3. The improvements, as seen by the darker blue areas, are consistent with the transformer's ability to model global dependencies specifically for wetlands (95), deciduous and mixed forests (42, 43), cultivated crops (82), and barren land (31).

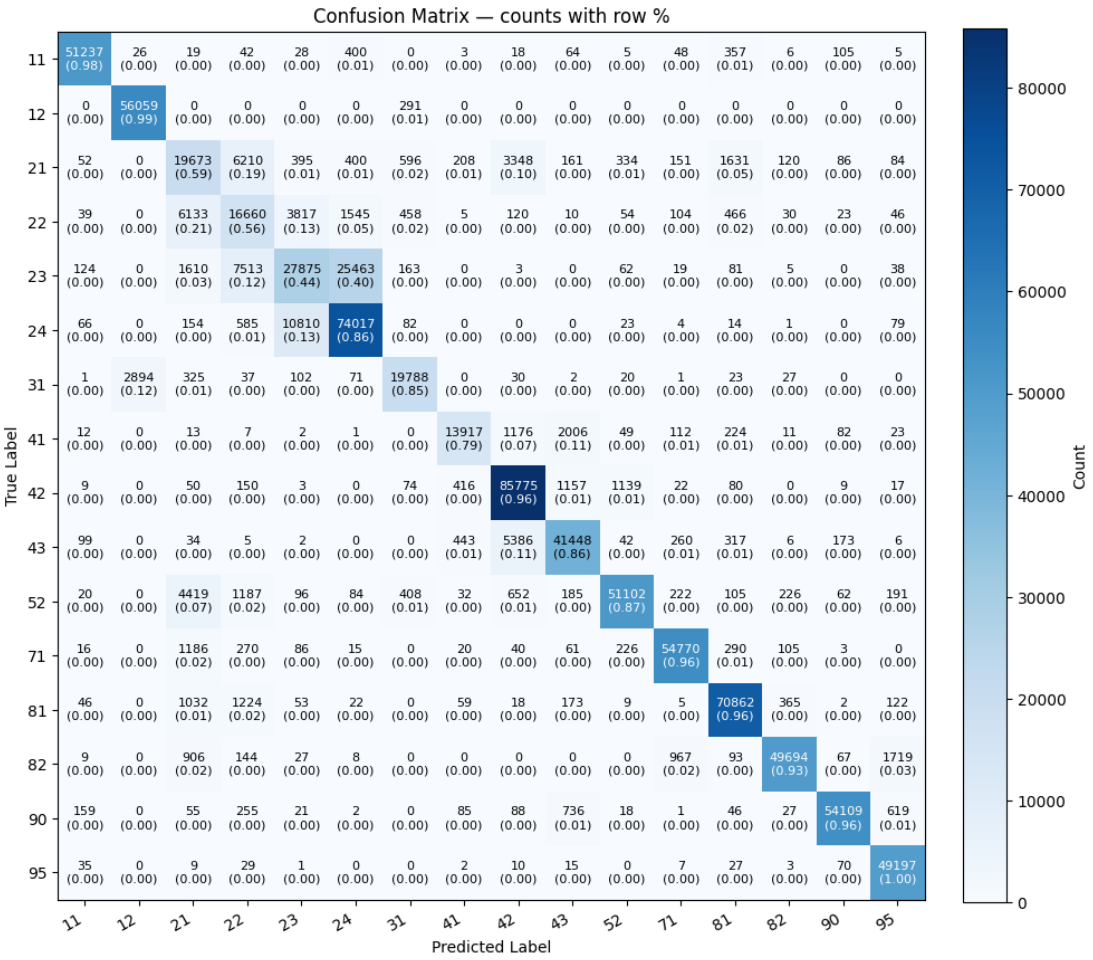

Figure 4.5.3a - SegFormer Confusion Matrix

Figure 4.5.3b demonstrates a smoother validation curve than the previous U-Net model, and a much lower separation between the train and validation losses. No signs of divergence as curves from all three metrics mirror each other, with validation being slightly rougher.

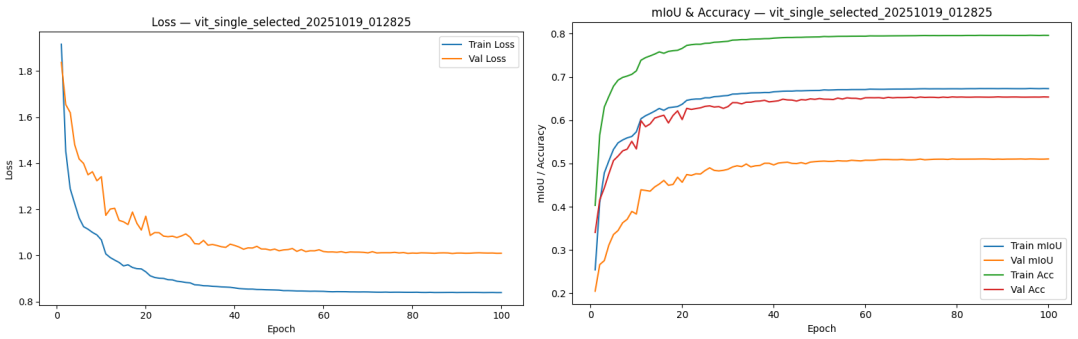

Figure 4.5.3b - SegFormer Loss and Accuracy

4.5.4 Hybrid CNN + Transformer

STransU2Net serves as the state-of-the-art Hybrid CNN + Transformer model and the last stage of experimentation for this praxis. This model was selected for its demonstrated ability to perform well on complex land use classification tasks due to its local feature extraction and global contextual attention. These two tasks are at the core of addressing many of the classification challenges detailed in §3.1.1. The STransU2Net architecture builds upon the U-Net foundation by adding Spatial Transformer Blocks (STB) within the U-Net's encoder path and Channel-Spatial Attention Blocks at the bottleneck. The U-Nets decoder up-sampling and skip-connections remain intact. StransU2Net offers the benefits from both ViT and U-Net architectures, which are shown in §4.5.2 and 4.5.3 to have comparable classification performance when each is used alone.

Table 4.5.4 STransU2Net Hyperparameter Search

	Explored	**Selected**
DIM	192,224,256,320	320
Heads	4,6,8,12,16	16
STB Window size	4,8,16	8
Learning Rate	1e-4,3e-3	0.0006
Dropout	0.05-0.25	0.1407
Label smoothing	0.00,0.15	0.0593
Batch size	64,96,112,128	112
StepLR step size	5,8,10,12,15	10
StepLR gamma	0.5,0.6,0.7,0.8	0.7
CSAB reduction r	8,12,16,24	16
BPB bottleneck ratio	2,4,8	4
W_Gamma	0.60,0.95	0.94

As shown in Table 4.5.4b below, STransU2Net achieved the highest performance of all the models tested. The model output supports hybrid transformers for accuracy gains (RQ3 and H3) with an Overall Accuracy of 92.7% and a 0.8170 mIoU. The loss of 2.2% accuracy from not including temporal features supports *Temporal Context for accuracy gains* (RQ4) research and supports the 2% accuracy gain target (H4).

Table 4.5.4b STransU2Net Temporal Results

	Overall Accuracy	**Balanced Accuracy**	**mIoU**
Including Temporal Features	92.7%	88.7%	0.8170
Not Including Temporal Features	90.5%	87.5%	0.7942

The confusion matrix in Table 4.5.4a below shows the strongest per-class accuracy out of all the models in this praxis, noted by dark average coloring on the diagonal. Urban classes remain troublesome, as mentioned in §4.5.1 and present through all of the models in this praxis, albeit improved by STransU2Net. The 0.8170 mIoU of the model supports this cleaner separation of borders. The overall pattern of the confusion matrix, when compared to the others in this praxis, indicates successful spatial precision.

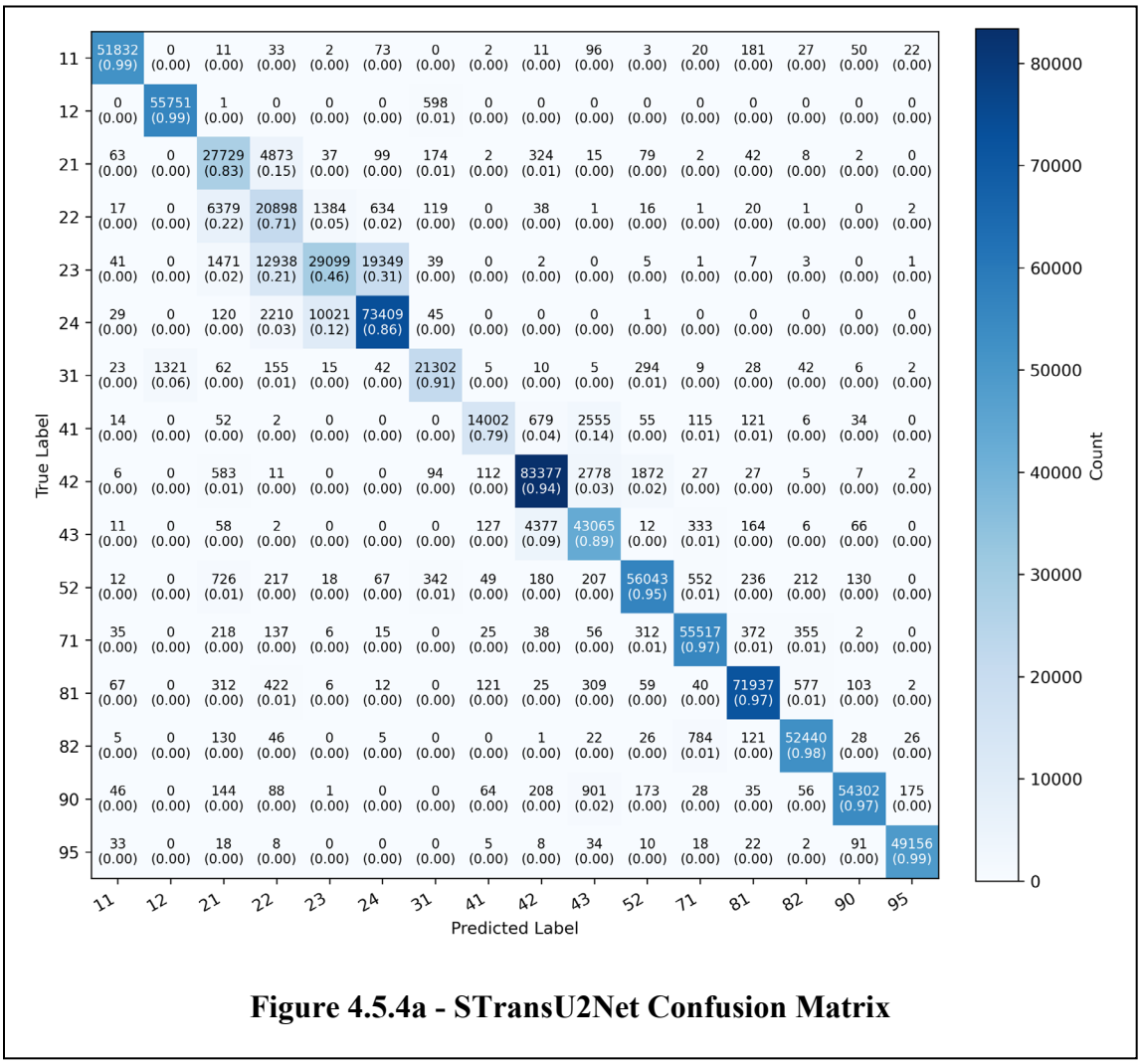

Figure 4.5.4a - STransU2Net Confusion Matrix

Figure 4.5.4b demonstrates the smoothest validation curve out of all the models, and a much lower separation between the train and validation losses.

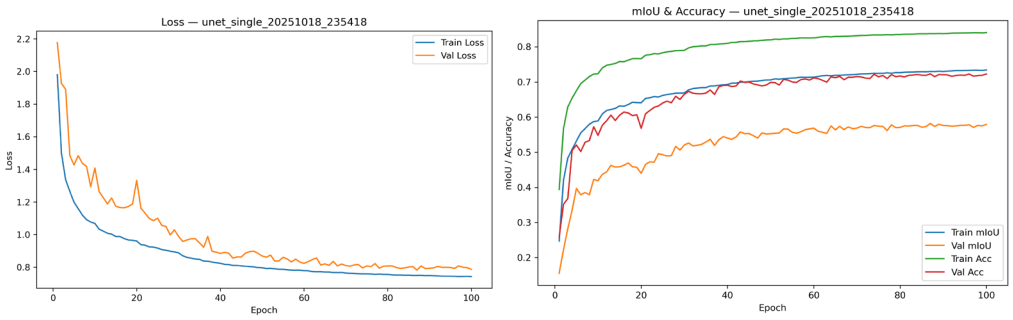

Figure 4.5.4b - STransU2Net Loss and Accuracy

4.6 Conclusion

This praxis applied a single pipeline to produce identical data splits for testing machine learning architectures in relation to LULC. Over this standardized dataset, deep neural networks with attention mechanisms outperformed both decision tree-based and U-Net models. In the overall research objective of developing a predictive model that achieves a 90% accuracy in land use classification, this praxis highlights the 92.7% overall accuracy achieved by the STransU2Net. Incorporating temporal components was shown to have a 2.2% increase in overall accuracy on the highest performing model. Praxis experimentation did validate accuracy gains from RF to U-Net and from U-Net to SegFormer, but at levels lower than suggested from a review of the published research literature. While not all hypotheses were supported in the experimentation, each led to essential contributions that provided incremental improvements toward the overall accuracy gain as described in Table 4.6 below.

Table 4.6 - Model Experiment Comparison

Model	Overall Accuracy	Balanced Accuracy	mIoU	Δ OA vs Previous	Δ mIoU vs Previous
Random Forest (RF)	78.6%	77.7%	0.6727	-	-
U-Net	87.6%	83.5%	0.7346	9	0.062
SegFormer (ViT)	88.5%	84.7%	0.7692	0.9	0.035
STransU2Net (Hybrid)	92.7%	88.7%	0.817	4.2	0.048

Chapter 5 – Discussion and Conclusions

This praxis set out to evaluate state-of-the-art machine learning models in the context of land use classification with the use of the National Land Cover Database (NLCD). The results of the experimentation from four research questions and hypotheses are recorded in the sections that follow. Overall, the praxis experiments validated the primary thesis statement.

Deep neural network models with attention mechanisms are required to improve NLCD land use classification.

Both of the models with attention mechanisms investigated for this praxis, the vision transformer (ViT) and the hybrid convolutional neural network (CNN) + transformer models, outperformed the U-Net and tree-based models that mirror the baseline models now used for the NLCD. This praxis met the overall research objectives of examining the limitations of current classification methods, investigating the role of attention mechanisms in model performance, and developing a predictive model that achieves a 90% accuracy in land use classification. The sections that follow discuss the research questions and hypotheses that serve as the foundation for meeting the research objectives.

5.1 Research Question and Hypothesis 1

RQ1: *Do deep neural networks outperform random forest models in differentiating forest types in the continental United States?*

Answer. *No.* U-Net achieved similar level 1 and level 2 accuracy scores as random forest.

H1: *Deep neural networks will outperform random forest models in accurately detecting forest types by 10% in the continental United States.*

Decision. *Not supported.* Deep neural network increases the detection of forest types by 0.6%.

In the comparison between deep neural networks and random forest models, a clear advantage is demonstrated in the shift to segmentation frameworks such as the U-Net used in this practice. Random forest (RF) baseline models achieved accuracy results in line with expectations from previous NLCD accuracy assessments. RF accuracy for level 1 forest is 97.67% with the most considerable confusion being 4% with urban development. U-Net accuracy for level 1 forest was also 97.22%. Improvements for U-Net are uneven across the 41/42/43 level 1 forest land use class square. Modest improvements for deciduous forest, and slight declines in both evergreen (42) and mixed forests (43), as detailed in Figure 5.1 below. (See Table A4 for NLCD Land Use Classes.)

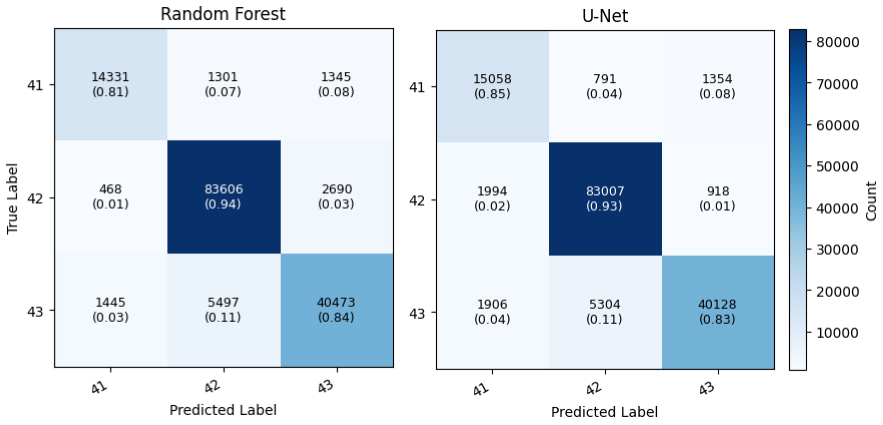

Figure 5.1 – Intra-Class Confusion for Forest Class

5.2 Research Question and Hypothesis 2

RQ2: *Do vision transformers outperform U-Net in LULC classification?*

Answer. *Yes.* SegFormer ViT outperformed U-Net in overall accuracy, balanced accuracy, and mean intersection over union (mIoU).

H2: *Vision transformers will outperform U-Net in LULC classification by 5% in accuracy.*

Decision. *Not Supported.* The SegFormer ViT model outperformed the U-Net by 0.9% overall accuracy and 1.2% balanced accuracy.

SegFormer ViT demonstrated a slight increase in classification accuracy over U-Net. However, the increase in accuracy did not reach the targeted 5% for statistical support of the hypothesis. Although only a 0.9% increase in overall accuracy was detailed, an increase in the mIoU of 0.035 was achieved. The increased mIoU indicates that the SegFormer has a clearer segmentation boundary quality. The accuracy results reflected class-level trade-offs, with U-Net achieving higher accuracies in five classes, and SegFormer achieving higher accuracies in seven. The SegFormer hierarchical architecture delivers different outcomes than those from a U-Net encoder-decoder. This shift in per-class accuracy, while not supportive of H2, suggests the potential of a hybrid architecture, which is addressed by H3. The most notable classes U-Net performed well on had features that are common across all regions: water (11), urban development (21, 23, 24), and grasslands (71). The most notable classes on which SegFormer performed well had features that change across regions: wetlands (95), deciduous and mixed forest (42, 43), cultivated crops (82), and barren land (31). SegFormer's global context and U-Nets' local feature recognition explain why H3 is supported more decisively than H2.

5.3 Research Question and Hypothesis 3

RQ3: *Do hybrid CNN + transformer models outperform U-Net in LULC classification?*

Answer. *Yes.* StransU2Net outperformed U-Net in overall accuracy, balanced accuracy, and mIoU.

H3: *Hybrid CNN + transformers will outperform U-Net in LULC classification by 5% in accuracy.*

Decision. *Supported.* STransU2Net outperformed U-Net by 5% Overall Accuracy.

Hybrid CNN + transformer models achieved the highest overall performance across all model types evaluated in this praxis. The STransU2Net hybrid model achieved better performance in every metric as seen in Table 5.3 below. The STransU2Net architecture added an attention mechanism that learns both local spatial and global spectral relationships. This dual-mechanism approach most notably improved difficult to classify minority intra-class relationships such as urban development classes, forest types, and wetland delineations. In these classes, the spectral and textural signatures overlap, reducing the effectiveness of the features created in this praxis for the rest of the models.

Table 5.3 - STransU2Net vs U-Net Performance

	Overall Accuracy	**Accuracy**	**Balanced Accuracy**	**mIoU**
U-Net	87.6%	85.3%	83.5%	0.7346
STransU2Net	92.7%	89.6%	88.7%	0.8170

5.4 Research Question and Hypothesis 4

RQ4: *Does incorporating temporal mechanisms reduce classification errors in land cover mapping in level 2 areas?*

Answer. *Yes.* Models with temporal features reduced errors relative to the baseline for that model.

H4: *Time-series analysis enhances the detection accuracy of LULC in both level 1 and level 2 classification by 2%.*

Decision. *Supported* at the 2% threshold by a positive effect averaging 4% across the different models tested.

Temporal features yielded positive results in every model except mIoU in the U-Net, as seen in Table 5.4 below. In the highest accuracy model, temporal features yielded a 2.2% increase in accuracy and a 0.0228 increase in mIoU. The accuracy increase was most notable in high-intensity urban development (24), forests (41, 42, 43), and shrubland (52). Five separate temporal features are among the top 20 most important features as measured by both Gini and Shapley Additive Explanations (SHAP). During both ablation studies conducted in U-Net and SegFormer, temporal surface reflectance was in the top 5 most essential features.

Table 5.4 - Temporal Accuracy Comparison Across Models

	RF	**U-Net**	**ViT**	**CNN + Transformer**
OA Gain	9.5%	1.6%	2.6%	2.2%
mIoU Gain	0.0454	-0.0132	0.0226	0.0228
Notable Classes	41,43,81			24,41,42,43,52

5.5 Contributions to Body of Knowledge

This praxis developed a comprehensive standardized feature engineering and testing for land-use and land-cover classification. This work operationalized a vast array of published features and techniques:

- Classical spectral indices like Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI).
- Evaluated newly published impervious surface indices like Band Ratio for Built-up Area (BRBA).
- Testing textural features derived from Gray Level Co-Occurrence Matrix (GLCM)
- Evaluated the use of One-hot encodings of Landsat categorical data.
- Integrated Environmental Protection Agency (EPA) Ecological Region (ecoregion) encodings.

By documenting and evaluating this comprehensive list of features, this praxis provides a reference point for future researchers.

This praxis also developed a phenology-first temporal featurization for national-scale land use and land cover mapping that utilized the MODIS MCD12Q2 dataset. The operational NLCD workflow is to synthesize leaf-on and leaf-off data based on a single target date. By deriving the leaf-on and leaf-off windows from the MODIS green-up, mid-green-up, mid-green-down, and dormancy dates, the seasonality is operated on recognized plant growth rather than the commonly used date-centric design. Within CONUS, the mid green-up date can vary by 10 months, a phenological targeted temporal feature that captures that variance with localized vegetation cycles.

Finally, this praxis benchmarked common machine learning architectures using industry-standard performance metrics, comparing decision trees, U-Net, Vision Transformers, and CNN + Transformers. Comparing these models on a standardized data pipeline isolated architectural differences from any data population or processing effects. This praxis successfully demonstrates a selection of a state-of-the-art national-scale LULC machine learning model for improved classification, surpassing the 90% overall accuracy target established for this research with the STransU2Net hybrid CNN + Transformer.

5.6 Recommendations for Future Research

During the course of this praxis, a potential area for future research developed that was not part of the original scope. Specifically, this considers refining the way quality assurance and other one-hot encoded layers are created as features. Currently, QA and EPA ecoregion data often consist of multiple discrete categories. Many of the hidden insights behind these categories exist within a common underlying spectrum. For example, cloud cover categories whether identified as cloud, cirrus, or dilated exist on a spectrum of cloud interference. Focusing on combining many of these categorical data into composites or gradient values would simplify the feature space and potentially improve model performance.

Another area for future research could be the extension of the windowed sampling procedure. Over 20 iterations were undertaken for this praxis, balancing class co-occurrence, geographic representation, ecological representation, and pixel-level data quality, demonstrating the complexity and impact of incorporating these steps into an operational workflow. A generalized sampling algorithm that optimizes these fields

would simplify data curation, reduce bias, and enable faster ideation and publication workflows for land-use and land-cover classification

References

Abdu, H. A. (2019). Classification accuracy and trend assessments of land cover- land use changes from principal components of land satellite images. *International Journal of Remote Sensing*, *40*(4), 1275–1300. https://doi.org/10.1080/01431161.2018.1524587

Abraham, N., & Khan, N. M. (2019, April). *A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation*. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). https://doi.org/10.1109/ISBI.2019.8759329

Adhikari, H., Heiskanen, J., Maeda, E. E., & Pellikka, P. K. E. (2015). DOES TOPOGRAPHIC NORMALIZATION OF LANDSAT IMAGES IMPROVE FRACTIONAL TREE COVER MAPPING IN TROPICAL MOUNTAINS? *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XL-7-W3*, 261–267. https://doi.org/10.5194/isprsarchives-XL-7-W3-261-2015

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). *Optuna: A Next-generation Hyperparameter Optimization Framework* (No. arXiv:1907.10902). arXiv. https://doi.org/10.48550/arXiv.1907.10902

Alexander, I. (2023). *Copyright and cartography: History, law, and the circulation of geographical knowledge*. Hart.

Anderson, J., Hardy, E., Roach, J., & Witmer, R. (1976). *A Land Use and Land Cover Classification System for Use with Remote Sensor Data* (Professional Paper) [Professional Paper]. USGS.

Arpitha, M., Ahmed, S. A., & Harishnaika, N. (2023). Land use and land cover classification using machine learning algorithms in google earth engine. *Earth Science Informatics*, *16*(4), 3057–3073. https://doi.org/10.1007/s12145-023-01073-w

Baig, M. H. A., Zhang, L., Shuai, T., & Tong, Q. (2014). Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. *Remote Sensing Letters*, *5*(5), 423–431. https://doi.org/10.1080/2150704X.2014.915434

Barnes, C. (2024, May). *Landsat Collection 2 Data Products and Characteristics*.

Barnes, C. A. (2014, November 17). *National Land Cover Database Evaluation, Visualization and Analysis (NLCD EVA) Tool*. ISPRS Technical Commission I Symposium, Denver, Colorado.

Beddar-Wiesing, S., Moallemy-Oureh, A., Kempkes, M., & Thomas, J. M. (2025). *Absolute Evaluation Measures for Machine Learning: A Survey* (No. arXiv:2507.03392). arXiv. https://doi.org/10.48550/arXiv.2507.03392

Benediktsson, J. A., Swain, P. H., & Ersoy, O. K. (1990). *Neural Network Approaches Versus Statistical Methods In Classification Of Multisource Remote Sensing Data*.

Berman, M., Triki, A. R., & Blaschko, M. B. (2018). The Lovasz-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-Over-Union Measure in Neural Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4413–4421. https://doi.org/10.1109/CVPR.2018.00464

Brown, J. (2024). *LSDS-2103 Annual National Land Cover Database (NLCD) Collection 1 Science Product User Guide -v1.0 2024_10_15*.

Burkov, A. (2025). *The hundred-page language models book*. True Positive Inc.

Campbell, J. B. (2011). *Introduction to remote sensing* (R. H. Wynne, Ed.; Fifth edition). The Guilford Press.

Campbell, J. B., Wynne, R. H., & Thomas, V. A. (2023). *Introduction to remote sensing* (6. edition). Guilford Press.

Campbell, J., & Wynne, R. (2011). *Introduction to Remote Sensing, fifth edition.* Guilford Publications.

Carreira, J., Caseiro, R., Batista, J., & Sminchisescu, C. (2012). Semantic Segmentation with Second-Order Pooling. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, & C. Schmid (Eds.), *Computer Vision – ECCV 2012* (Vol. 7578, pp. 430–443). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33786-4_32

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), 6. https://doi.org/10.1186/s12864-019-6413-7

Clark, A., Phinn, S., & Scarth, P. (2023). Optimised U-Net for Land Use–Land Cover Classification Using Aerial Photography. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, *91*(2), 125–147. https://doi.org/10.1007/s41064-023-00233-3

Committee on the Future of Water Quality in Coeur d'Alene Lake, Water Science and Technology Board, Division on Earth and Life Studies, & National Academies of Sciences, Engineering, and Medicine. (2022). *The Future of Water Quality in Coeur d'Alene Lake* (p. 26620). National Academies Press. https://doi.org/10.17226/26620

de Andrade, A. (2019). *Best Practices for Convolutional Neural Networks Applied to Object Recognition in Images*.

Dekker, B., Ton, B., Meijer, J., Bouali, N., Linssen, J., & Ahmed, F. (2023). Point Cloud Analysis of Railway Infrastructure: A Systematic Literature Review. *IEEE Access*, *11*, 134355–134373. https://doi.org/10.1109/ACCESS.2023.3337049

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* (No. arXiv:2010.11929). arXiv. https://doi.org/10.48550/arXiv.2010.11929

EROS Center. (2024, February). *NLCD_metric_spotlight*. U.S. Geological Survey. https://www.usgs.gov/media/images/30-years-national-land-cover-database-nlcd

Fernández-Armesto, F. (2011). *Human Geography*. 738–741.

Freudenberg, M., Nölke, N., Agostini, A., Urban, K., Wörgötter, F., & Kleinn, C. (2019). Large Scale Palm Tree Detection in High Resolution Satellite Images Using U-Net. *Remote Sensing*, *11*(3), 312. https://doi.org/10.3390/rs11030312

Fu, Y., Zhu, Z., Liu, L., Zhan, W., He, T., Shen, H., Zhao, J., Liu, Y., Zhang, H., Liu, Z., Xue, Y., & Ao, Z. (2024). Remote Sensing Time Series Analysis: A Review of Data and Applications. *Journal of Remote Sensing*, *4*, 0285. https://doi.org/10.34133/remotesensing.0285

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, *31*(14), 2225–2236. https://doi.org/10.1016/j.patrec.2010.03.014

Gerlach, A., Anderson, J., Park, A., Hardy, E., Tomlinson, R. F., & Wray, J. (1971, June 28). *A land use classification scheme for use with remote sensor data*. Inter-Agency Steering Committee on Land Use Information and Classification.

Gou, X., Liao, C., Zhou, J., Ye, F., & Lin, Y. (2024). *FIF-UNet: An Efficient UNet Using Feature Interaction and Fusion for Medical Image Segmentation* (No. arXiv:2409.05324). arXiv. https://doi.org/10.48550/arXiv.2409.05324

Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: An Overview* (No. arXiv:2008.05756). arXiv. https://doi.org/10.48550/arXiv.2008.05756

Gray, J., Sulla-Menashe, D., & Friedl, M. A. (2019). User Guide to Collection 6.1 MODIS Land Cover Dynamics (MCD12Q2) Product. *NASA's Earth Observator*.

Gu, Y., Wylie, B., Boyte, S., Picotte, J., Howard, D., Smith, K., & Nelson, K. (2016). An Optimal Sample Data Usage Strategy to Minimize Overfitting and Underfitting Effects in Regression Tree Models Based on Remotely-Sensed Data. *Remote Sensing*, *8*(11), 943. https://doi.org/10.3390/rs8110943

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., & Tao, D. (2023). A Survey on Vision Transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 87–110. https://doi.org/10.1109/TPAMI.2022.3152247

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Heidari, M., Kazerouni, A., Soltany, M., Azad, R., Aghdam, E. K., Cohen-Adad, J., & Merhof, D. (2023, January 9). HiFormer: Hierarchical Multi-scale Representations Using Transformers for Medical Image Segmentation. *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. https://doi.org/10.1109/WACV56688.2023.00614

Hemati, M., Hasanlou, M., Mahdianpari, M., & Mohammadimanesh, F. (2021). A Systematic Review of Landsat Data for Change Detection Applications: 50 Years of Monitoring the Earth. *Remote Sensing*, *13*(15), 2869. https://doi.org/10.3390/rs13152869

Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *Journal of Forestry Research*, *32*(1), 1–6. https://doi.org/10.1007/s11676-020-01155-1

Jin, S., Dewitz, J., Danielson, P., Granneman, B., Costello, C., Smith, K., & Zhu, Z. (2023). National Land Cover Database 2019: A New Strategy for Creating Clean Leaf-On and Leaf-Off Landsat Composite Images. *Journal of Remote Sensing*, *3*, 0022. https://doi.org/10.34133/remotesensing.0022

Jin, S., Homer, C., Yang, L., Danielson, P., Dewitz, J., Li, C., Zhu, Z., Xian, G., & Howard, D. (2019). Overall Methodology Design for the United States National Land Cover Database 2016 Products. *Remote Sensing*, *11*(24), 2971. https://doi.org/10.3390/rs11242971

Karakizi, C., Karantzalos, K., Vakalopoulou, M., & Antoniou, G. (2018). *Detailed Land Cover Mapping from Multitemporal Landsat-8 Data of Different Cloud Cover*.

Kayhan, O. S., & Gemert, J. C. van. (2020). On Translation Invariance in CNNs: Convolutional Layers can Exploit Absolute Spatial Location. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14262–14273. https://doi.org/10.1109/CVPR42600.2020.01428

Khan, M., Hanan, A., Kenzhebay, M., Gazzea, M., & Arghandeh, R. (2024). Transformer-based land use and land cover classification with explainability using satellite imagery. *Scientific Reports*, *14*(1), 16744. https://doi.org/10.1038/s41598-024-67186-4

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022). Transformers in Vision: A Survey. *ACM Computing Surveys*, *54*(10s), 1–41. https://doi.org/10.1145/3505244

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Lee, J., Han, D., Shin, M., Im, J., Lee, J., & Quackenbush, L. J. (2020). Different Spectral Domain Transformation for Land Cover Classification Using Convolutional Neural Networks with Multi-Temporal Satellite Imagery. *Remote Sensing*, *12*(7), 1097. https://doi.org/10.3390/rs12071097

Letard, M., Collin, A., Corpetti, T., Lague, D., Pastol, Y., & Ekelund, A. (2022). *Classification of Land-Water Continuum Habitats Using Exclusively Airborne Topobathymetric Lidar Green Waveforms and Infrared Intensity Point Clouds*.

Limin Yang, Homer, C., Hegge, K., Chengquan Huang, Wylie, B., & Reed, B. (2001). A Landsat 7 scene selection strategy for a national land cover database. *IGARSS*

2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217), *3*, 1123–1125. https://doi.org/10.1109/IGARSS.2001.976766

Ling, X., Zhu, Y., Ming, D., Chen, Y., Zhang, L., & Du, T. (2022). Feature Engineering of Geohazard Susceptibility Analysis Based on the Random Forest Algorithm: Taking Tianshui City, Gansu Province, as an Example. *Remote Sensing*, *14*(22), 5658. https://doi.org/10.3390/rs14225658

Liu, G., Diao, K., Zhu, J., Wang, Q., & Li, M. (2024). STransU2Net: Transformer based hybrid model for building segmentation in detailed satellite imagery. *PLOS ONE*, *19*(9), e0299732. https://doi.org/10.1371/journal.pone.0299732

Liu, S., & Gu, G. (2017). Improving the Impervious Surface Estimation from Hyperspectral Images Using a Spectral-Spatial Feature Sparse Representation and Post-Processing Approach. *Remote Sensing*, *9*(5), 456. https://doi.org/10.3390/rs9050456

Liu, W., Tian, T., Xu, W., Yang, H., Pan, X., Yan, S., & Wang, L. (2022). *PHTrans: Parallelly Aggregating Global and Local Representations for Medical Image Segmentation* (No. arXiv:2203.04568). arXiv. https://doi.org/10.48550/arXiv.2203.04568

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965

Lundberg, S., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, *30*.

https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

McRoberts, J., & Lynch, J. (1972). *Earth Resources Technology Satellites-A (ERTS-A): Press Kit [Press release]*. National Aeronautics and Space Administration.

Orhan, A. E., & Pitkow, X. (2024). *Skip Connections Eliminate Singularities* (No. arXiv:1701.09175). arXiv. https://doi.org/10.48550/arXiv.1701.09175

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sensing*, *12*(14), 2291. https://doi.org/10.3390/rs12142291

Pisanty, I., Gauthier, D., Hannah, L., Hirvonen, H., Marshall, I., & Wiken. (1997). *Ecological regions of North America: Toward a common perspective*. The Commission for Environmental Cooperation.

Rawat, K. S., Kumar, S., & Garg, N. (2024). Statistical comparison of simple and machine learning based land use and land cover classification algorithms: A case study. *Journal of Water Management Modeling*. https://doi.org/10.14796/JWMM.H524

Röchner, P., Marques, H. O., Campello, R. J. G. B., Zimek, A., & Rothlauf, F. (2025). *Robust Statistical Scaling of Outlier Scores: Improving the Quality of Outlier Probabilities for Outliers (Extended Version)* (Vol. 15268, pp. 215–222). https://doi.org/10.1007/978-3-031-75823-2_18

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-*

Assisted Intervention – MICCAI 2015, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Sabr, H. A. (2024). Integrate Terrain Variables and RapidEye Satellite in Vegetation Indices, for Identifying Forest Cover Area and Density: A Case Study in Mountainous Iraqi Kurdistan Region (IKR). *Passer Journal of Basic and Applied Sciences*, *6*(1), 107–114. https://doi.org/10.24271/psr.2023.408530.1353

Sayler, K. (2022, February). *Landsat 9 Data Users Handbook*. Department of the Interior U.S. Geological Survey.

Shapiro, L. G., & Stockman, G. C. (2001). *Computer vision*. Prentice Hall.

Sievers, M. (2014). *Land Use Loading Literature Review Task Summary and Results*. Tetra Tech.

Sim, W.-D., Yim, J.-S., & Lee, J.-S. (2024). Assessing Land Cover Classification Accuracy: Variations in Dataset Combinations and Deep Learning Models. *Remote Sensing*, *16*(14), 2623. https://doi.org/10.3390/rs16142623

Socher, R., Huval, B., Bath, B., Manning, C. D., & Ng, A. Y. (2012). Convolutional-Recursive Deep Learning for 3D Object Classification. *Advances in Neural Information Processing Systems*, *25*. https://proceedings.neurips.cc/paper/2012/hash/3eae62bba9ddf64f69d49dc48e2dd214-Abstract.html

Stehman, S. V., Wickham, J. D., Smith, J. H., & Yang, L. (2003). Thematic accuracy of the 1992 National Land-Cover Data for the eastern United States: Statistical methodology and regional results. *Remote Sensing of Environment*, *86*(4), 500–516. https://doi.org/10.1016/S0034-4257(03)00128-7

Swapna, M., Sharma, Y. K., & Prasad, B. (2020). CNN Architectures: Alex Net, Le Net, VGG, Google Net, Res Net. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(6), 953–959. https://doi.org/10.35940/ijrte.F9532.038620

Tanaka, S., Kawamura, K., Maki, M., Muramoto, Y., Yoshida, K., & Akiyama, T. (2015). Spectral Index for Quantifying Leaf Area Index of Winter Wheat by Field Hyperspectral Measurements: A Case Study in Gifu Prefecture, Central Japan. *Remote Sensing*, *7*(5), 5329–5346. https://doi.org/10.3390/rs70505329

Udall, S. (1966). *Statement by Secretary of the Interior Stewart L. Udall on Earth photography from Gemini 4 [Press release]*. U.S. Department of the Interior. https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/atoms/files/1966_Udall.pdf

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Vogelmann, J. (2001). *Completion of the 1990s National Land Cover Data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources*.

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, *2018*, 1–13. https://doi.org/10.1155/2018/7068349

Warren, L. L. (2022, June). *Intalco SO 2 Attainment Plan Modeling Report by AECOM, June 2022*. AECOM.

Wickham, J., Homer, C., Vogelmann, J., McKerrow, A., Mueller, R., Herold, N., & Coulston, J. (2014). The Multi-Resolution Land Characteristics (MRLC) Consortium—20 Years of Development and Integration of USA National Land Cover Data. *Remote Sensing*, *6*(8), 7424–7441. https://doi.org/10.3390/rs6087424

Wickham, J., Stehman, S. V., Gass, L., Dewitz, J. A., Sorenson, D. G., Granneman, B. J., Poss, R. V., & Baer, L. A. (2017). Thematic accuracy assessment of the 2011 National Land Cover Database (NLCD). *Remote Sensing of Environment*, *191*, 328–341. https://doi.org/10.1016/j.rse.2016.12.026

Wickham, J., Stehman, S. V., Sorenson, D. G., Gass, L., & Dewitz, J. A. (2023). Thematic accuracy assessment of the NLCD 2019 land cover for the conterminous United States. *GIScience & Remote Sensing*, *60*(1), 2181143. https://doi.org/10.1080/15481603.2023.2181143

Wunderle, S., Gross, T., & Hüsler, F. (2016). Snow Extent Variability in Lesotho Derived from MODIS Data (2000–2014). *Remote Sensing*, *8*(6), 448. https://doi.org/10.3390/rs8060448

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *Advances in Neural Information Processing Systems*, *34*, 12077–12090. https://proceedings.neurips.cc/paper/2021/hash/64f1f27bf1b4ec22924fd0acb550c235-Abstract.html

Yang, L., Driscol, J., Sarigai, S., Wu, Q., Lippitt, C. D., & Morgan, M. (2022). Towards Synoptic Water Monitoring Systems: A Review of AI Methods for Automating Water Body Detection and Water Quality Monitoring Using Remote Sensing. *Sensors*, *22*(6), 2416. https://doi.org/10.3390/s22062416

Yao, J., Zhang, B., Li, C., Hong, D., & Chanussot, J. (2023). Extended Vision Transformer (ExViT) for Land Use and Land Cover Classification: A Multimodal Deep Learning Framework. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–15. https://doi.org/10.1109/TGRS.2023.3284671

Yazıcı, Z. A., Öksüz, İ., & Ekenel, H. K. (2024). GLIMS: Attention-Guided Lightweight Multi-Scale Hybrid Network for Volumetric Semantic Segmentation. *Image and Vision Computing*, *146*, 105055. https://doi.org/10.1016/j.imavis.2024.105055

Zhang, L., Zhang, L., & Du, B. (2016). Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geoscience and Remote Sensing Magazine*, *4*(2), 22–40. https://doi.org/10.1109/MGRS.2016.2540798

Zhang, X., A., M., Tan, B., D., M., & Yu, Y. (2012). Long-Term Detection of Global Vegetation Phenology from Satellite Instruments. In X. Zhang (Ed.), *Phenology and Climate Change*. InTech. https://doi.org/10.5772/39197

Zhang, Y., Liu, H., & Hu, Q. (2021, July 10). TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. https://doi.org/10.1007/978-3-030-87193-2_2

Zhao, H., Ma, Y., Chen, F., Liu, J., Jiang, L., Yao, W., & Yang, J. (2018). Monitoring Quarry Area with Landsat Long Time-Series for Socioeconomic Study. *Remote Sensing*, *10*(4), 517. https://doi.org/10.3390/rs10040517

Zhao, X., Wu, J., Wang, H., Gao, X., & Zhao, L. (2022). Injecting spectral indices to transferable convolutional neural network under imbalanced and noisy labels for Landsat image classification. *International Journal of Digital Earth*, *15*(1), 437–462. https://doi.org/10.1080/17538947.2022.2036833

Appendix

Table A1 – QA_PIXEL Bit Values

Bit(s)	Flag	Description	Values
0	Fill	Pixel is image data or fill	0 = image data; 1 = fill data
1	Dilated Cloud	Cloud dilation indicator	0 = not dilated / no cloud; 1 = dilated cloud
2	Cirrus	Cirrus confidence (single-bit)	0 = none/low confidence; 1 = high confidence cirrus
3	Cloud	Cloud confidence (single-bit)	0 = not high; 1 = high
4	Cloud Shadow	Cloud-shadow confidence (single-bit)	0 = not high; 1 = high
5	Snow	Snow/ice confidence (single-bit)	0 = not high; 1 = high
6	Clear	Clear condition	0 = cloud or dilated cloud set; 1 = neither cloud nor dilated cloud set
7	Water	Water indicator	0 = land/cloud; 1 = water
8–9	Cloud Confidence	Cloud confidence (2-bit)	00 = none set; 01 = low; 10 = medium; 11 = high
10–11	Cloud Shadow Confidence	Cloud-shadow confidence (2-bit)	00 = none set; 01 = low; 10 = reserved; 11 = high
12–13	Snow/Ice Confidence	Snow/ice confidence (2-bit)	00 = none set; 01 = low; 10 = reserved; 11 = high

14–15	Cirrus Confidence	Cirrus confidence (2-bit)	00 = none set; 01 = low; 10 = reserved; 11 = high

Table A2 – QA_RADSAT Bit Values

Bit(s)	Flag	Description	Values
0	Band 1 Data Saturation	Radiometric saturation status for Band 1	0 = no saturation; 1 = saturated data
1	Band 2 Data Saturation	Radiometric saturation status for Band 2	0 = no saturation; 1 = saturated data
2	Band 3 Data Saturation	Radiometric saturation status for Band 3	0 = no saturation; 1 = saturated data
3	Band 4 Data Saturation	Radiometric saturation status for Band 4	0 = no saturation; 1 = saturated data
4	Band 5 Data Saturation	Radiometric saturation status for Band 5	0 = no saturation; 1 = saturated data
5	Band 6 Data Saturation	Radiometric saturation status for Band 6	0 = no saturation; 1 = saturated data
6	Band 7 Data Saturation	Radiometric saturation status for Band 7	0 = no saturation; 1 = saturated data
7	Unused	Reserved / not used	0 = not checked
8	Band 9 Data Saturation	Radiometric saturation status for Band 9	0 = no saturation; 1 = saturated data
9	Unused	Reserved / not used	0 = unused
10	Unused	Reserved / not used	0 = unused

11	Terrain Occlusion	Line-of-sight blocked by terrain	0 = no terrain occlusion; 1 = terrain occlusion
12	Unused	Reserved / not used	0 = unused
13	Unused	Reserved / not used	0 = unused
14	Unused	Reserved / not used	0 = unused
15	Unused	Reserved / not used	0 = unused

Table A3 – QA_AEROSOL Bit Values

Bit(s)	Flag	Description	Values
0	Fill	Pixel fill indicator	0 = pixel is not fill; 1 = pixel is fill
1	Valid Aerosol Retrieval	Retrieval validity	0 = not valid; 1 = valid
2	Water	Water indicator	0 = not water; 1 = water
3	Unused	Reserved / not used	0 = unused
4	Unused	Reserved / not used	0 = unused
5	Interpolated Aerosol	Aerosol value was interpolated	0 = not interpolated; 1 = interpolated
6–7	Aerosol Level	Aerosol level (2-bit field)	00 = climatology; 01 = low; 10 = medium; 11 = high

Table A4 – NLCD Land Use Classes

Class	Description
10	Water
11	Open Water: Areas of open water, generally with less than 25% cover of vegetation or soil.
12	Perennial Ice/Snow: Areas characterized by a perennial cover of ice and/or snow, generally greater than 25% of total cover
20	Developed
21	Developed, Open Space: areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20% of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes.
22	Developed, Low Intensity – areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20% to 49% percent of total cover. These areas most commonly include single-family housing units.
23	Developed, Medium Intensity – areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50% to 79% of the total cover. These areas most commonly include single-family housing units.
24	Developed, High Intensity – highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses and commercial/industrial. Impervious surfaces account for 80% to 100% of the total cover.
30	Barren
31	Barren Land (Rock/Sand/Clay) – areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover.
40	Forest
41	Deciduous Forest – areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75% of the tree species shed foliage simultaneously in response to seasonal change.
42	Evergreen Forest – areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75% of the tree species maintain their leaves all year. Canopy is never without green foliage.
43	Mixed Forest – areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. Neither deciduous nor evergreen species are greater than 75% of total tree cover.
50	Shrubland

52	Shrub/Scrub – areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20% of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions.
70	Herbaceous
71	Grassland/Herbaceous – areas dominated by graminoid or herbaceous vegetation, generally greater than 80% of total vegetation. These areas are not subject to intensive management such as tilling but can be utilized for grazing.
80	Planted/Cultivated
81	Pasture/Hay – areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20% of total vegetation.
82	Cultivated Crops – areas used to produce annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and perennial woody crops such as orchards and vineyards. Crop vegetation accounts for greater than 20% of total vegetation. This class also includes all land being actively tilled.
90	Wetlands
90	Woody Wetlands – areas where forest or shrubland vegetation accounts for greater than 20% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.
95	Emergent Herbaceous Wetlands – areas where perennial herbaceous vegetation accounts for greater than 80% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

Table A5 – 2019 NLCD Accuracy Assessment

Class	Count	Percent Total Pixels	Accuracy
11	150	4.62%	97%
12	4	0.12%	50%
21	163	5.02%	72%
22	133	4.10%	39%
23	186	5.73%	47%
24	127	3.91%	63%
31	156	4.81%	64%
41	198	6.10%	78%
42	337	10.39%	89%
43	115	3.54%	63%
52	421	12.97%	80%
71	454	13.99%	72%
81	117	3.61%	70%
82	302	9.31%	91%
90	175	5.39%	54%
95	207	6.38%	54%