Dual LSTM Pipelines for Multi-Temporal Larch Plantation Mapping in the Changqing Region: A Supervised and Self-Supervised Approach to Data-Scarce Mountainous Forests

by Josh Manto

Signature Work Product, in partial fulfillment of the Duke Kunshan University Undergraduate Degree Program

March 7, 2025

Signature Work Program
Duke Kunshan University

APPROVALS

Mentor: Ding Ma, Division of Natural and Applied Sciences
Co-mentor: Binbin Li, Division of Natural and Applied Sciences
Paul Stanley, Associate Dean of Undergraduate Studies

Abstract

Forests are highly dynamic ecosystems, and understanding their species composition at regional scales is critical for biodiversity conservation and land management. Here, we present an LSTM-based methodology for classifying dominant tree species in the Changqing region using multi-temporal Sentinel-2 observations. First, we employed a custom JavaScript script in Google Earth Engine (GEE) to extract monthly spectral bands and vegetation indices (e.g., NDVI, EVI, NDWI) over a 12-month period, capturing key phenological signals. Next, we curated groundtruth data from field surveys, cleaning and consolidating labeled points into a standardized format for modeling. We then trained a Long Short-Term Memory (LSTM) neural network to learn the temporal signatures of different tree species, using time-series inputs that spanned 12 monthly observations of each pixel sampled on a 30-day revisit time frame. Our LSTM framework was validated on 2,900+ groundtruth points which was obtained from fieldwork in the region during the years 2022 and 2023; the groundtruth contained data on locations of Chinese larch, Japanese larch, coniferous, broadleaf, and mixed forest types in the region. We finally applied the trained model to unseen GEE patches, generating a high-resolution species distribution map of the Changqing region. The resulting map reveals spatial patterns of larch, deciduous broadleaf, and other forest classes, offering insights into biodiversity mapping and potential areas of conservation priority.

Acknowledgements

I would like to thank my signature work mentor, Dr. Ding Ma, and Co-mentor: Dr. Binbin Li, as well as the signature work office at Duke Kunshan University for their support throughout my research process.

LIST OF TABLES AND FIGURES

LIST OF FIGURES

1 INTRODUCTION

1.1 Problem Statement

The rapid expansion of human activities has significantly altered global land use and land cover, posing substantial threats to biodiversity worldwide (Newbold et al., 2015). China, in particular, has undertaken extensive afforestation and reforestation efforts, resulting in the largest area of tree plantations globally (FAO, 2020). These plantations are primarily established as a means of forest restoration and to meet timber demands. However, the monoculture nature of these plantations often leads to simplified spatial structures, making them susceptible to disease outbreaks and causing instability in community and ecosystem functions (Bremer & Farley, 2010). Consequently, plantations are sometimes referred to as "green deserts" due to their low biodiversity compared to natural forests (Lindenmayer & Hobbs, 2004).

In the southern Qinling Mountains, the Changqing region serves as a critical biodiversity hotspot, home to endangered species such as the giant panda (Ailuropoda melanoleuca) and the golden snub-nosed monkey (Rhinopithecus roxellana) (Loucks et al., 2001). These species rely on mixed coniferous and broad-leaved forests that provide diverse food sources and habitats, including abundant bamboo understory essential for the giant panda (Hull et al., 2014; Tuanmu et al., 2013). From the late 1960s through the 1990s, intensive deforestation and subsequent reforestation activities replaced large areas of natural forest with larch plantations, primarily Japanese larch *(Larix kaempferi*) and Chinese larch (*Larix gmelinii var. principis-rupprechtii*), due to their fast growth and economic value (Zhang, 2023).

These changes have altered the landscape structure, impacting local ecosystems and the habitats of endangered species. Moreover, selective logging, which is important for managing plantations and improving biodiversity, is no longer legal inside nature reserves due to environmental regulations (State Forestry Administration of China, 2017). Therefore, understanding how landscape dynamics, such as plantation shape and age, influence biodiversity is crucial for conservation efforts.

In his study, Zhang (2023) aimed to map the current distribution of larch plantations and detect historical deforestation activities in the southern Qinling region using multi-temporal remote sensing imagery. Zhang used supervised classification techniques, specificallty random forest, on satellite images, with the help of previous fieldworks conducted by his team in the region to collect ground truth data—including 113 points in Huangbaiyuan National Nature Reserve (NNR), 96 in Changqing NNR, and 90 in Guanyinshan NNR—Zhang was able to reveal the landscape patterns of larch plantations and analyze their correlation with historical deforestation. However, his methods were labor-intensive, requiring significant fieldwork for data collection and manual tagging, not to mention the region's complex and mountainous terrain.

Additionally, Zhang's methodology faced challenges due to shadows in the satellite images caused by the rugged topography and temporal changes in vegetation phenology, which affected classification accuracy. The accuracy assessment of Zhang's current distribution map showed an overall accuracy of 76.1% and a Kappa coefficient of 0.654, indicating substantial agreement but also room for improvement. These limitations highlight the need for more efficient and accurate methods for mapping vegetation types in such challenging environments.

1.2 LSTM-Based Time-Series Classification as a Solution

Early efforts to map tree plantations in the Changqing region frequently relied on supervised classification methods (e.g., random forests) applied to single-date satellite images (Zhang, 2023). While these approaches offer a snapshot of forest cover, they are limited in their ability to capture seasonal changes in vegetation reflectance. In mountainous areas—where topographic shadows and cloud contamination are common—single-date imagery often suffers from spectral ambiguities that reduce classification accuracy. Specifically, random forest classifiers trained on one-off acquisitions can misinterpret shadowed pixels as different forest types, or fail to differentiate larch stands with overlapping spectral signatures when viewed at a single point in time.

Time-series analysis provides a more robust way to account for these complexities. By observing forest pixels across multiple months, it becomes possible to track phenological patterns that vary among species. For example, Chinese larch (Larix gmelinii var. principis-rupprechtii) and Japanese larch (Larix kaempferi) exhibit distinct leaf-on and leaf-off periods, typically flushing needles in spring (April–May) and shedding them in late autumn (October–November), whereas evergreen conifers retain needles year-round (Bremer & Farley, 2010; Zhang, 2023). These seasonal shifts in canopy properties produce unique temporal signatures in vegetation indices (e.g., NDVI) and spectral bands.

To leverage these multi-temporal signals, we adopted a Long Short-Term Memory (LSTM) framework. LSTM networks, a class of recurrent neural networks, excel at modeling sequential dependencies—here, the month-to-month evolution of spectral and vegetation index values. By providing approximately 12 months of observations for each pixel, the LSTM can learn how reflectance patterns rise and fall in response to leaf emergence, peak greenness, and senescence. This temporal perspective mitigates errors introduced by shadows or clouds in any single scene, as the network infers class membership from the broader trend rather than a single-date spectral snapshot. That means larch plantations, natural broadleaf forests, and mixed coniferous stands can each be uniquely characterized based on their time-series profiles, achieving more accurate species-level classification across diverse terrain.

1.3 Objectives of the Study

1. **Map the entire Changqing region** to locate and classify all tree species, including larch plantations and other tree species using supervised machine learning.
2. **Benchmark the classification results** by comparing them with existing models and datasets, assessing the accuracy and performance of the self-supervised approach against traditional supervised methods.

1.4 Vegetation Types to Be Classified

- **Larch Species:**
    - larch_JP: Japanese larch *(Larix kaempferi)*
    - larch_CN: Chinese larch *(Larix gmelinii var. principis-rupprechtii)*
- **Forest Types:**
    - deci_broad: deciduous broadleaf forest
    - ever_broad: evergreen broadleaf forest
    - ever_coni: Evergreen coniferous forest
    - mix_coni_broad: Mixed coniferous and broad-leaved forest
- **Other Land Cover:**
    - Shrubland

1.5 Literature Review

Remote sensing has long been recognized as a crucial tool for mapping and monitoring forest plantations, particularly in regions with rugged terrain and dynamic phenological patterns (Jones & Vaughan, 2010). Recent advancements in satellite imagery—exemplified by the European Space Agency's Sentinel-2 Mission—have enabled finer spatial and temporal resolution than was previously possible (Immitzer, Atzberger, & Koukal, 2016; Persson, Lindberg, & Reese, 2018). These improvements are important for capturing the seasonal variations of larch plantations, whose spectral signatures differ markedly between needle-on and needle-off phases (Bremer & Farley, 2010; Zhang, 2023).

1.5.1 Differences in Plantation Patterns and the Need for Multi-Temporal Data

Zhang (2023) found that larch plantations in the southern Qinling region vary greatly in size and distribution: Huangbaiyuan National Nature Reserve (NNR) features one large continuous patch, whereas Changqing NNR presents a more heterogeneous mosaic of plantation sizes. Guanyinshan NNR, by contrast, is characterized by small, highly fragmented larch stands. Capturing these spatial nuances often requires multi-temporal imagery. Single-date images run the risk of misclassifying terrain shadows and

overlooks the phenological cues critical for distinguishing larch from other coniferous and broadleaf species (Persson et al., 2018). Multi-temporal Sentinel-2 imagery thus provides “phenological fingerprints,” allowing species to be recognized from their leaf-on, peak, and senescence phases.

1.5.2 Key Spectral Regions: Red-Edge, SWIR, and Visible Bands

Numerous studies underscore the value of red-edge (around 705–783 nm) and shortwave-infrared (SWIR, 1610–2190 nm) bands for tree species classification (Immitzer et al., 2016; Persson et al., 2018). The red-edge region captures chlorophyll content and subtle shifts in canopy condition, while SWIR reflectance responds to foliage water content and structural compounds (e.g., lignin, cellulose). These properties make it possible to discriminate larch from evergreen conifers and broadleaf species, especially when observed over multiple points in the growing season (Zhang, 2023).

The visible blue (around 490 nm) and green (560 nm) bands also offer complementary insights into canopy structure. Young plantations, which often have more exposed soil or sparse understory, reflect more strongly in the blue band, whereas denser canopies exhibit lower reflectance due to greater shadowing and leaf layering (Immitzer et al., 2016). Healthy foliage further elevates green-band reflectance, helping to distinguish broadleaf forests from needleleaf stands.

1.5.3 Phenological Signatures in Larch Plantations

A major advantage of multi-temporal data lies in capturing phenological transitions—an especially pronounced phenomenon in larch species, which lose needles in winter and regain them in spring. Larch stands typically exhibit higher near-infrared (NIR) reflectance in summer when needles are fully flush, then experience increased red reflectance and lower NIR signatures as needles turn yellow and drop in late autumn (Bremer & Farley, 2010; Persson et al., 2018). By contrast, evergreen conifers retain their needles year-round and thus maintain relatively stable red and NIR signatures. Incorporating imagery from spring, mid-summer, and autumn can markedly improve classification accuracy in mixed forests, where larch, evergreen conifers, and deciduous broadleaf species co-occur (Immitzer et al., 2016).

1.5.4 Importance of Indices for Topographically Complex Areas

The mountainous topography of the Qinling region introduces significant variation in illumination, which can lead to confusion between shadows and true spectral signals. Normalized difference indices (NDIs), such as the classic NDVI (Tucker, 1979) or moisture-oriented indices combining NIR and SWIR, can help normalize brightness values and mitigate topographic shadow effects (Persson et al., 2018). By focusing on ratios, these indices are less sensitive to absolute reflectance differences, thereby offering more consistent performance across slopes and aspects.

1.5.5 Implications for Larch Plantation Mapping

The literature so far demonstrates that multi-temporal approaches are central to accurately mapping plantations in mountainous regions (Persson et al., 2018; Zhang, 2023). Time-series Sentinel-2 observations can capture larch phenology—bud flush, senescence, and needle drop—and thus separate larch from visually similar coniferous species. Furthermore, the use of red-edge and SWIR bands, along with vegetation indices, and NDIs, all aid in distinguishing plantation types and detecting sub-canopy variations in moisture and structure. Hence, we employ all these in the data extraction part.

1.5.6 Preliminary Methods for Mapping Tree Species (Non-Spatiotemporal)

Early efforts in forest cover and species mapping often employed classifiers that relied on single-date or limited multi-spectral data without explicitly modeling temporal variation. Traditional approaches such as the Maximum Likelihood Classifier (MLC) and Minimum Distance Classifier were widely used due to their straightforward implementation and relatively low computational costs (Jones & Vaughan, 2010). Although these methods can perform adequately in regions with clear spectral separability, they struggle to account for seasonal changes—particularly problematic in mountainous terrain where vegetation phenology and illumination angles vary substantially (Immitzer et al., 2016).

Subsequent research introduced machine learning algorithms like Support Vector Machines (SVM) and Random Forests (RF) as more robust alternatives for single-date

imagery classification. SVMs maximize the margin between classes in feature space, reducing misclassification in scenarios with overlapping spectral signatures (Persson et al., 2018). Random Forests, which aggregate decisions from multiple decision trees, have demonstrated strong accuracy for forest mapping owing to their resilience to noise and flexibility across diverse feature sets (Zhang, 2023). These approaches, however, still do not explicitly account for time-series dynamics—making them vulnerable to confusion between species whose reflectance values converge at certain times of the year.

1.5.7 Advanced Spatiotemporal Approaches

Recognizing the limitations of single-date classification, more recent studies emphasize spatiotemporal or strictly temporal methods that harness multi-date satellite observations. One line of research employs Time-Weighted Dynamic Time Warping (TWDTW), a technique originally used to match similar patterns in time-series data by allowing non-linear alignments (Maus et al., 2016). TWDTW is especially beneficial when vegetation phenology is slightly out of phase due to differences in elevation or microclimates, as it warps the time axis to optimally match spectral signatures.

Another family of methods uses harmonic modeling (also referred to as harmonic regression or harmonic analysis) to decompose time-series data into seasonal and trend components. By fitting sinusoidal functions to indices like NDVI or EVI, harmonic methods capture periodic changes that can help separate deciduous species, such as larch, from evergreens (Zhang et al., 2019). These approaches have proven effective in forest disturbance monitoring, where sudden deviations from harmonic patterns often indicate logging or disease outbreaks.

Meanwhile, supervised deep learning models—such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks—have gained traction for spatiotemporal remote sensing. LSTMs, in particular, excel at modeling sequential dependencies and can handle irregular gaps or cloud-affected imagery (Hochreiter & Schmidhuber, 1997). By learning phenological “fingerprints” unique to each species, LSTM-based classifiers typically outperform single-date RF or SVM approaches,

especially in complex terrains like the Qinling Mountains (Zhang, 2023). Nevertheless, these methods often require extensive labeled data for training—an obstacle in remote, biodiverse regions where ground surveys are logistically difficult.

1.5.8 Unsupervised Tree Species Mapping

While supervised learning has dominated tree species classification, unsupervised or self-supervised approaches remain comparatively rare in forestry applications. Traditional unsupervised clustering algorithms like K-Means or Hierarchical Clustering treat spectral values as unlabeled features and group pixels based on similarity (Jones & Vaughan, 2010). However, simple clustering methods often conflate species with overlapping reflectance properties—especially when phenology is not taken into account or when spectral differences are subtle (Immitzer et al., 2016).

More recent developments involve autoencoder architectures and Bayesian clustering algorithms that leverage unlabeled time-series data to discover naturally emerging classes. For instance, an LSTM autoencoder can learn compressed representations of spectral-temporal sequences, preserving key phenological cues without requiring external labels (Hochreiter & Schmidhuber, 1997). A Bayesian Gaussian Mixture Model (BGMM) can then partition these latent features into clusters, indicating potential species groups or subtypes (Maus et al., 2016). Such self-supervised frameworks are especially promising in data-scarce regions: they reduce the reliance on large ground-truth datasets and may reveal previously unrecognized vegetation categories. Although these unsupervised methods are still emerging in operational forest mapping, preliminary studies suggest they can complement or even rival supervised classifiers—particularly where temporal signals strongly differentiate species (Zhang, 2023).

1.5.9 Balancing Supervised and Unsupervised Strategies for Limited Ground Truth Data

Recent work by Huang et al. (2023) illustrates a spectral–temporal deep learning framework built on Transformer architectures to classify plantation forests in two Chinese regions, Mengjiagang and Gaofeng. Their approach self-adapts to extract

latent features from a dense 5-day Sentinel-2 time series and achieves high mapping accuracy, largely thanks to extensive ground-truth data—118,454 reference points in Mengjiagang and 312,976 in Gaofeng. This robust labeled dataset allowed their supervised method to generalize effectively, leveraging Transformers' capacity to capture nuanced temporal signals.

By contrast, the current study faces a significantly smaller labeled dataset of roughly 4,631 ground-truth points in the Changqing region. Relying solely on a supervised model—such as an LSTM framework—would likely constrain classification performance and risk overfitting, given the high dimensionality of 55 features across 12 monthly observations. Indeed, while Huang et al. (2023) also employ a spectral–temporal strategy to handle similarly rich feature sets, their large annotated inventory offsets the complexity of model training. Here, a fully supervised workflow is inhibited by limited sample diversity and label coverage.

Consequently, this research adopts both supervised and unsupervised (self-supervised) pipelines to maximize information extraction from time-series Sentinel-2 data. The supervised LSTM model benefits from the labeled examples that do exist, learning phenological signatures for key forest types. Simultaneously, an unsupervised or autoencoder-based method leverages the unlabeled portions of the dataset, clustering latent representations without requiring exhaustive field observations. This dual design, similar in spirit to Huang et al.'s (2023) feature extraction but replacing the Transformer backbone with an LSTM encoder–decoder, better suits the smaller ground-truth inventory available in the Changqing region. The result is a more flexible framework that balances label scarcity against the inherent complexity of multi-temporal remote sensing data.

2. Methodology

2.1 Study Area and Sentinel-2 Time Series Data

Our study region covers an area of approximately 183.5 km², defined by a 1792 × 1024 pixel grid (each pixel at 10 m spatial resolution). For efficient processing, we divided this

region into 28 equally sized patches, each measuring 256 × 256 pixels (65,536 pixels per patch). Consistent with prior work highlighting the importance of multi-temporal data in capturing phenological signals (Persson et al., 2018; Zhang, 2023), we employed a custom JavaScript script within Google Earth Engine (GEE) to compile monthly Sentinel-2 observations over a full 12-month cycle.

We extracted both spectral bands (B1–B12) and vegetation indices such as NDVI, EVI, and NDWI at 10 m resolution. All data was processed to surface reflectance and extracted from the Harmonized Sentinel-2 MSI dataset. The built-in QA60 cloud-masking band was applied, which selectively retained cloud-free scenes during key phenological periods (e.g., spring leaf flush, summer peak foliage, and autumn senescence). Each pixel's spectral and index values—along with latitude, longitude, and sampling date—were recorded in a CSV for each of the 12 monthly revisit periods, yielding a total of 786,432 data points. Table 1 provides an overview of the Sentinel-2 MultiSpectral Instrument (MSI) features included in our time series.

We then collected ground-truth data through field surveys conducted in 2022 and 2023. This resulted in 4,631 reference points labeled for dominant forest classes: two larch species (Larix gmelinii var. principis-rupprechtii and Larix kaempferi), broadleaf (deciduous and evergreen), coniferous, mixed conifer–broadleaf, and shrubland (Table 2). After cleaning and normalizing these samples, we matched each reference point's 12-month time series to the extracted Sentinel-2 features. Given that larch species exhibit distinct leaf-on and leaf-off periods (Bremer & Farley, 2010; Zhang, 2023), assembling a yearlong temporal profile was critical for revealing monthly nuances in reflectance and vegetation indices.

Name	**Scale**	**Wavelength**	**Description**
B1	60 m	443.9 nm (S2A)	Aerosols

B2-4	10 m	496.6 nm (S2A), 560 nm (S2A), 560 nm (S2A)	Blue, Green, Red
B5-7	20 m	703.9 nm (S2A), 740.2 nm (S2A), 782.5 nm (S2A)	Red Edge 1, 2, 3
B8	10 m	835.1 nm (S2A)	NIR
B8A	20 m	864.8 nm (S2A)	Red Edge 4
B9	60 m	945 nm (S2A)	Water Vapor
B11-12	20 m	1613.7 nm (S2A), 2202.4 nm (S2A)	SWIR 1, SWIR 2
AOT	0.001		Aerosol Optical Thickness
WVP	0.001		Water Vapor Pressure
SCL			Scene Classification Map (mask for no data)
TCI_R, G, B			True Color Image, Red, Green, Blue channel
MSK_CLDPRB			Cloud Probability Map

Table 1: Details of Sentinel-2 MSI data

Class name	**Tree species**	**Sample count**
larch_CN	Chinese larch	230

larch_JP	Japanese larch	1340
deci_broad	Decidious broadleaf	1212
ever_broad	Evergreen broadleaf	40 (ignore this class, not a lot in souhern cq)
ever_coni	Evergreen coniferous	820
mix_coni_broad	Mix coniferous broadleaf	789
Shrubland	Shrubland	200

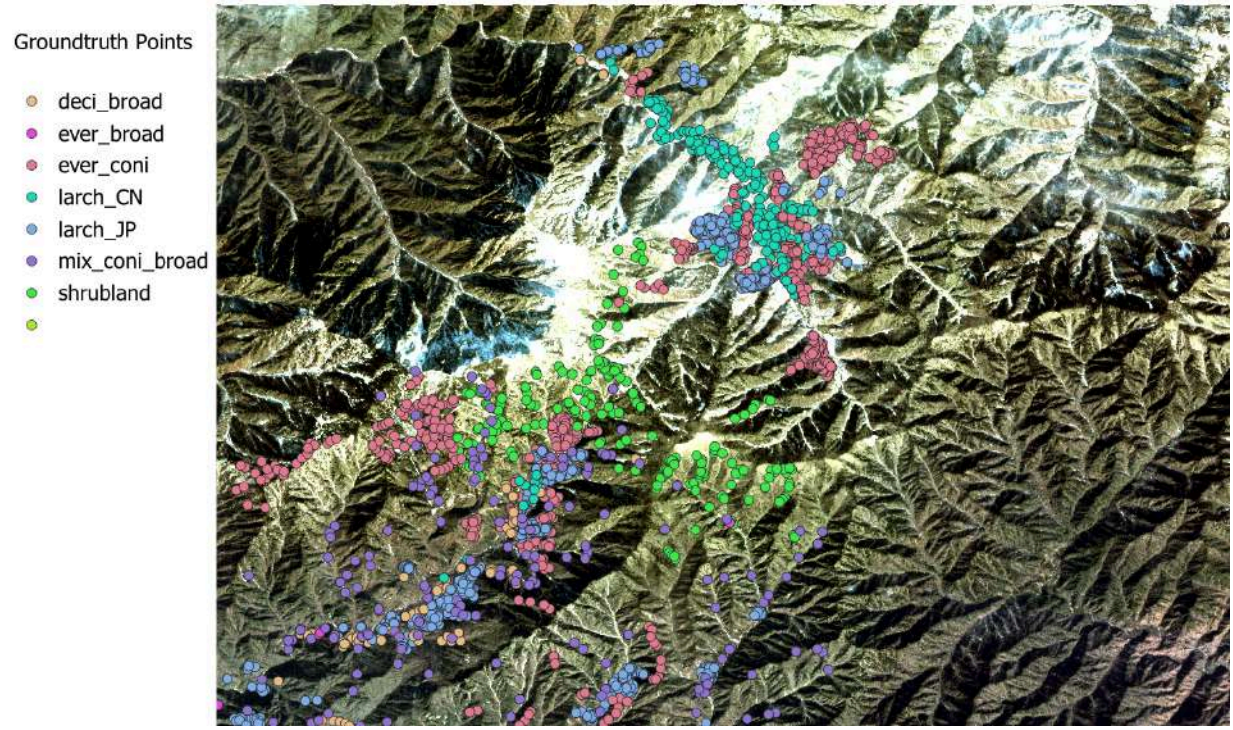

Table 2: Ground-truth Data for Dominant Classes.

2.2 Full Pipeline

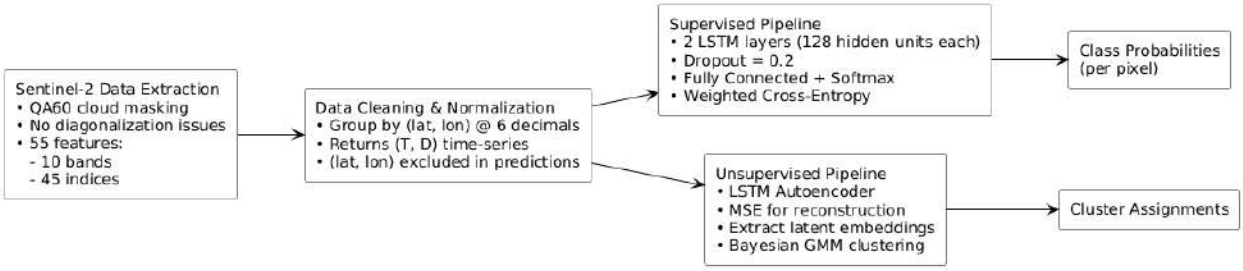

Figure 1: Full Pipeline Architecture

The end-to-end workflow (Fig. 1) begins with monthly Sentinel-2 data extraction and QA60 cloud masking within Google Earth Engine (GEE). Next, we group pixels by latitude-longitude (rounded to six decimals) and organize each into a consistent 12-month time series of 55 features (including spectral bands and derived indices). We then branch into two parallel pipelines—a Supervised LSTM classifier for labeled data and an Unsupervised LSTM autoencoder for unlabeled data—each producing distinct outputs that either categorize the pixels (class probabilities) or reveal naturally emerging clusters, respectively. This dual approach combines the benefits of label-based training for known classes with self-supervised discovery of latent spectral-temporal patterns, thereby addressing data scarcity while capturing crucial phenological signals.

2.3 LSTM architecture

To handle the high-dimensional data (55 total features × 12 time steps) and mitigate challenges associated with limited ground-truth samples, we implemented a two-branch LSTM approach (Table 3). This design follows recommendations in the literature for integrating spatiotemporal signals—particularly phenological patterns—into classification models (Hochreiter & Schmidhuber, 1997; Persson et al., 2018).

- **Supervised Branch.** We trained a bidirectional LSTM classifier on the 4,631 labeled points. Each input sequence corresponds to one pixel's 12-month spectral and index values, capturing seasonal shifts such as larch needle flush or broadleaf senescence. The LSTM contains two layers of 128 hidden units, with a softmax output for five major forest categories (larch_CN, larch_JP, ever_broad,

deci_broad, ever_coni, plus classes for mixed stands and shrubland). Because the distribution of sample counts is unbalanced (e.g., only 40 evergreen broadleaf points), we applied a weighted cross-entropy loss to reduce bias toward larger classes.

- **Self-Supervised Branch.** In parallel, an LSTM autoencoder compresses the same 12-month sequences into a 64-dimensional latent representation. By reconstructing the input (via a mean squared error objective), the autoencoder preserves core phenological traits—such as dips in reflectance during leaf-off—without requiring labels. We then applied a Bayesian Gaussian Mixture (BGM) clustering algorithm to these latent codes, allowing emergent groups to form based on subtle spectral-temporal differences. This step can reveal, for instance, nuanced splits in conifer stands or understory conditions not clearly distinguished by the limited ground-truth data.

Pipeline	Data	Model	Architecture	Optimization	Output
Supervised	Labeled 12-month time-series data	Bidirectional LSTM Classifier	- 2 LSTM layers, each with 128 hidden units - Dropout probability = 0.2 - Fully connected layer for classification (softmax)	Weighted cross-entropy (to handle class imbalance)	Class probabilities (per pixel / sample)
Unsupervised	Unlabeled 12-month time-series data	LSTM Autoencoder + Bayesian GM Clustering	- 2 LSTM encoder layers, each with 128 hidden units - Latent dimension = 64 - 2 LSTM decoder layers mirroring the encoder - Dropout probability = 0.2	Mean squared error (sequence reconstruction)	Cluster assignments (e.g., discovered sub-classes)

Table 3: LSTM Pipelines and Model Architecture

This dual pipeline addresses the scarcity of labeled data—underscored in our literature review (see Sections 2.8 and 2.9)—by capitalizing on both supervised learning (for classes we can confidently label) and unsupervised learning (to discover structure beyond our limited ground-truth). The parallel usage of an LSTM architecture in both branches ensures that key temporal signals are modeled consistently, building upon prior research that demonstrates LSTM's effectiveness in mountainous, phenologically dynamic regions (Zhang, 2023).

3. Results and Analysis

3.1 Validating Unsupervised Results

Epoch	Matched GT points	ARI	NMI	Purity
300	2,892	0.3291	0.3372	0.6591
700	2,941	0.0409	0.1901	0.4947
Clusters & Their Dominant Groundtruth Classes (300 Epochs)				
Cluster 0: majority larch_JP (44%) Cluster 3: majority mix_coni_broad (69%) Cluster 4: majority larch_JP (39%) Cluster 5: majority larch_CN (100%) Cluster 6: majority ever_coni (78%) Cluster 7: majority larch_JP (66%) Cluster 8: majority ever_coni (99.8%) Cluster 9: majority ever_coni (49%)				
Clusters & Their Dominant Groundtruth Classes (700 Epochs)				
Cluster 0: majority larch_JP (35%) Cluster 3: majority larch_JP (72%) Cluster 4: majority ever_coni (75%)				

Cluster 5: majority mix_coni_broad (35%)
Cluster 6: majority larch_JP (43%)
Cluster 7: majority mix_coni_broad (56%)
Cluster 8: majority mix_coni_broad (35%)
Cluster 9: majority ever_coni (99%)

Table 4: Unsupervised Classification results

To evaluate the clustering consistency of our self-supervised pipeline, we compared the LSTM autoencoder + Bayesian Gaussian Mixture (BGM) model's cluster assignments against approximately 2,900 groundtruth samples that fell within the inference bounding box. We computed three primary metrics—Purity, Adjusted Rand Index (ARI), and Normalized Mutual Information (NMI)—at two different training epochs (Table 4).

- Purity measures the proportion of a cluster's points that belong to its most common groundtruth class. Formally, it is computed by assigning each cluster to the majority groundtruth class within it and calculating the overall fraction of correctly assigned samples. In simpler terms, Purity tells us how "pure" each cluster is—whether most of the points in a given cluster actually belong to the same real-world category (e.g., a vegetation type). A higher Purity means fewer mislabeled points in each cluster.

- Adjusted Rand Index (ARI) quantifies the similarity between the predicted clustering and the groundtruth labels, adjusting for chance. It evaluates how well the clusters align with the actual categories by considering both correct and incorrect pairwise assignments. Put simply, ARI checks whether points that should be grouped together are actually clustered correctly, and whether unrelated points are kept apart. An ARI of 1 means perfect clustering, while 0 indicates clustering no better than random chance.

- Normalized Mutual Information (NMI) captures the mutual dependence between cluster labels and groundtruth classes. It is derived from information theory and

measures how much knowing the predicted cluster label reduces uncertainty about the true class label. In layman's terms, NMI tells us how much the discovered clusters "inform" us about the real categories—if NMI is high, it means the clusters strongly correspond to actual vegetation types, whereas a low NMI suggests that the clusters are not meaningful in relation to the groundtruth labels.

Higher values across these metrics indicate a stronger agreement between our model's discovered clusters and the known vegetation types. See that table 4 provides a side-by-side summary of the two epochs. We first observed a sharp contrast in performance: at epoch 300, Purity is nearly 0.66, ARI is around 0.33, and NMI is about 0.34, whereas at epoch 700, Purity drops below 0.50, ARI plunges to ~0.04, and NMI to ~0.19. This decline suggests that the autoencoder's extended optimization favored a latent structure less correlated with broad species labels, possibly reflecting fine-scale reflectance or stand-age differences rather than distinct forest types. Meaning there could have been significant overfitting as we increased the epoch size. Similar observations have been reported by AlZuhair et al. (2023), who noted that over-training deep clustering models may lead to "memorizing" noise rather than salient class boundaries, which is very common for a lot of machine learning models. In Huang et al.'s paper (2023) they also talk about how dealing with a lot of features, such as in our context of 55 features, there is a risk of having to deal with the "curse of dimensionality" which are challenges of working with high-dimensional data where distance metrics and clustering algorithms become less effective because points tend to become "equidistant" - meaning that as the number of dimensions increases, the differences between data points become less pronounced, making it harder to distinguish meaningful clusters.

3.1.1 Comparison to Other Studies and Dataset Constraints

One factor that likely hinders the clustering model's ability to align with labeled classes is the limited availability of groundtruth data—about 2,900 points in a broad

mountainous forest region. In contrast, studies such as Huang et al. (2023) assembled much larger field inventories (over 100,000 points each in Mengjiagang and Gaofeng plantations) and Wang et al. (2022) similarly collected thousands of training and test samples (e.g., ~20,000 larch, ~11,000 Mongolian pine, etc.). With such extensive ground inventories, typical supervised or semi-supervised models often achieve higher accuracies because they can more reliably learn class decision boundaries. Our unsupervised approach, however, has no direct label guidance and must discover patterns from limited samples—thus reducing the achievable Purity and ARI.

3.1.2 Discussion of Segmentation Mapping

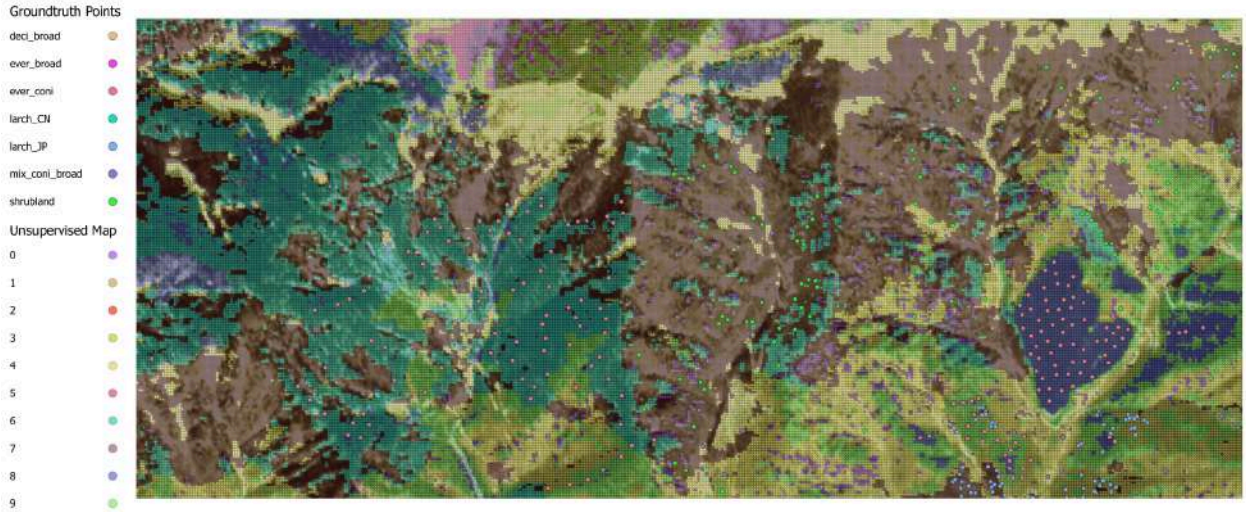

Figure 2: Unsupervised Mapping Overlayed with Groundtruth Points

Despite these constraints, Figure 2 highlights the level of detail our LSTM autoencoder + Bayesian Gaussian Mixture (BGM) model can produce in challenging terrain. Notably, at the center-right portion of the map, the ever_coni points (shown as groundtruth in that region) are grouped into a single coherent cluster that closely matches the labeled data. This demonstrates the model's ability to isolate specific vegetation types (evergreen conifer) even when other surrounding classes—such as *larch_CN*, *mix_coni_broad*, or *shrubland*—are nearby. From visual inspection, each local terrain patch also exhibits distinct boundaries that align well with topographic variations (e.g., ridge vs. valley areas). While some boundary "bleeding" is inevitable in complex

mountainous environments, it was still able to delineate zones that correspond to groundtruth vegetation classes.

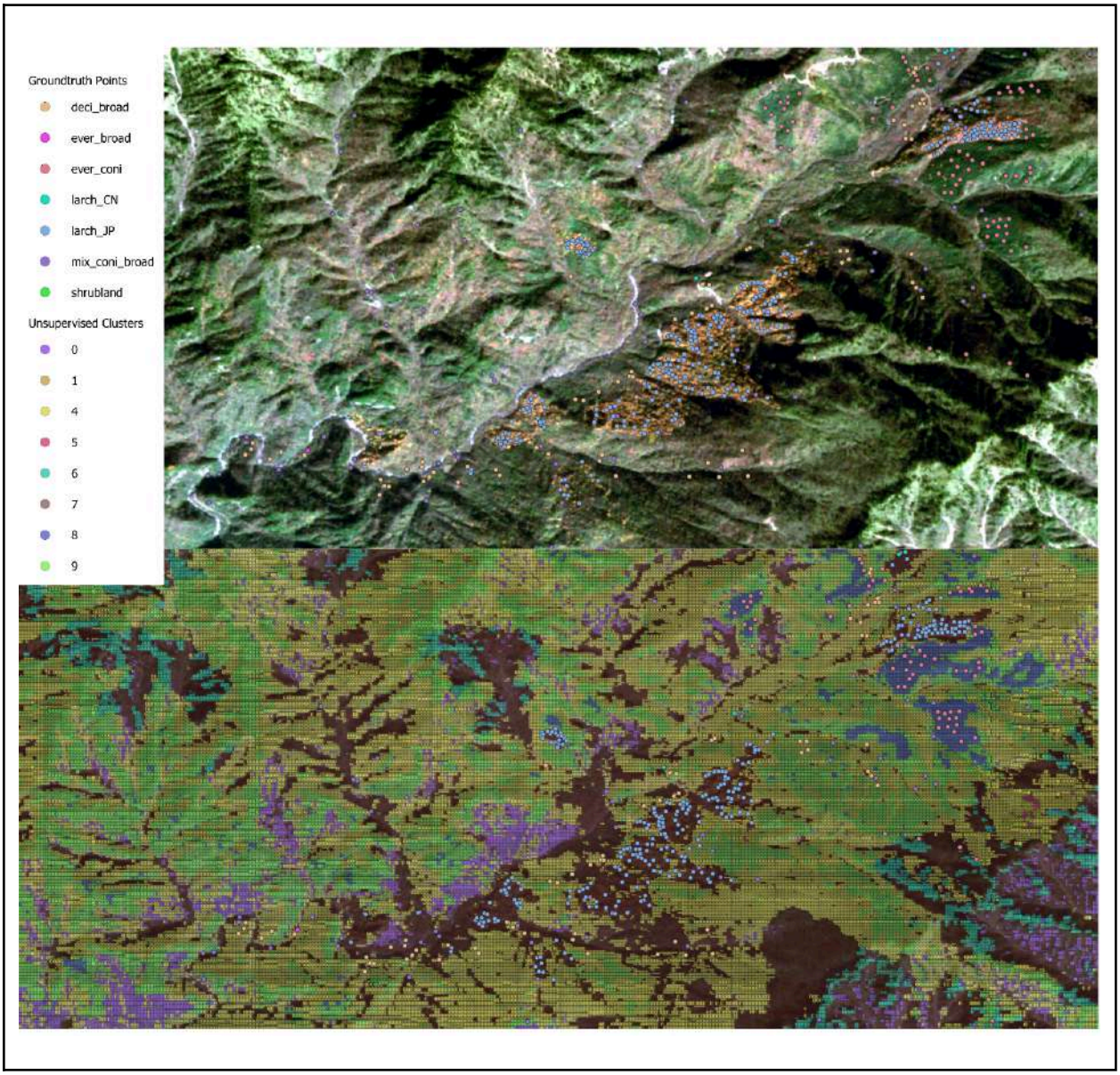

Figure 3: Groundtruth Japanese Larch Plantations and Unsupervised Segmentation Results

Moving to Figures 3 and 4, we see a closer look at Japanese Larch (larch_JP) plantations in autumn.

- Figure 3 shows a PlanetScope image taken on October 29, 2019 where these larch plantations can be clearly observed turning gold, reflecting the characteristic senescence stage in late fall. With the naked eye, one might guess

there are perhaps two or three main color groupings—“golden trees,” “dark green/evergreen patches,” and “bare or shrub” land—yet the unsupervised model finds a more nuanced segmentation.

- It also demonstrates the resulting cluster assignments over the same area, revealing that our pipeline can detect *multiple* distinct vegetation clusters within what appears uniform to the untrained observer. This outcome indicates that phenological cues—like the timing of leaf coloration and abscission—are being leveraged effectively by the LSTM autoencoder, yielding fine-grained separations among “golden” larch stands, darker green conifer stands, and potentially mixed transitional zones.

These results contrast with typical segmentation tasks in remote sensing literature that frequently focus on cropland mapping, where fields often have clean, artificial boundaries and drastically different spectral signatures between crop types. In mountainous forests, however, subtle transitions between stands, understory composition, and shading from uneven terrain make the segmentation task far more difficult. Shadows present a particular problem: if the sun azimuth angle is low or partially obstructed by steep slopes, large areas of the image can be shadowed or underexposed, leading to misclassified patches. These geometric complications are amplified in hilly or mountainous topographies. The model appears capable of separating *ever_coni* and *larch_JP* stands in shaded, slope-facing corners, suggesting a certain degree of robustness to illumination changes. However, we lack detailed groundtruth specifically in these heavily shadowed areas, so further validation would be required to confirm the accuracy of those classifications under such conditions.

Nevertheless, the trade-off is that continuing training to 700 epochs allowed the model to “memorize” subtle spectral fluctuations or noise rather than focusing on large-scale class distinctions, thereby reducing Purity, ARI, and NMI. Despite the inherent difficulty of separating multiple conifer and broadleaf classes across rugged terrain with minimal groundtruth, the epoch 300 outcome still showcases impressive spatial detail. The unsupervised approach manages to approximate the actual forest mosaics, capturing variation driven by phenology, species composition, and local topography in a way that

simpler models—which rely on straightforward, well-defined boundaries—cannot easily replicate. While our best Purity (~0.66) does not rival supervised benchmarks from richer inventory datasets, it still indicates that unsupervised methods can reveal meaningful structure in a data-limited, mountainous forest environment.

3.2 Supervised Classification Results

To gauge our LSTM classifier's effectiveness across all 28 patches, we merged each patch's predictions with the groundtruth data and derived a combined confusion matrix. Table 5 condenses five key metrics from this global matrix, providing insight into overall accuracy, class balance, and model robustness. When examined at the patch level, results show that some patches reached perfect (100%) or near-perfect (>98%) accuracy, while others fell to 0%—in most cases, because only a single groundtruth class occupied that patch, yet the model consistently assigned a different label.

Metric	Value
Overall Accuracy	0.9445
Macro-F1	0.7676
Cohen's Kappa	0.9125
Balanced Accuracy	0.7753
Weighted-F1	0.9291

Table 5: *Aggregated metrics for the supervised LSTM classifier across all 28 patches.*

	deci_broad	ever_coni	larch_CN	larch_JP	shrubland

deci_broad	12	224	174	3462	84
ever_coni	100	35610	20	430	100
larch_CN	20	20	9400	880	20
larch_JP	38	109	700	56163	90
shrubland	0	10	0	0	9010

Table 6: *Global Confusion Matrix for the Supervised LSTM classifier across all 28 patches.*

Although we originally collected around 4,800 labeled “groundtruth points,” each field reference typically corresponds to **multiple underlying 10 m pixels**—especially when mapped onto our 28 sub-patches. In other words, every labeled coordinate or polygon can produce dozens (or even hundreds) of pixel-level matches in the inference grid. Hence, when we merge these labels with the model predictions, the total number of matched groundtruth–prediction pairs expands to **116,676** in the combined confusion matrix. This explains why the final validation dataset is much larger than the raw count of groundtruth points.

From Table 5 and Table 6, we note an overall accuracy of 94.45%, indicating that nearly 95% of the 116,676 matched points were classified correctly. This strong performance aligns well with regions featuring abundant groundtruth data, such as central Huangbaiyuan, where large, contiguous stands of *larch_JP* and *ever_coni* are accurately mapped. Meanwhile, a Balanced Accuracy of about 0.7753—lower than the raw accuracy—reflects persistent difficulties in classes with fewer samples (e.g., certain

broadleaf or shrubland types). A salient limitation is the model's tendency to overestimate *larch_JP* in areas lacking robust groundtruth, notably patch 0_0 (west Northern Changqing) and patch 3_0 (east Northern Changqing), both showing 0% accuracy in Table 7.

Patch 0_0 confusion matrix: **Northern Changqing Region** Accuracy=0.0000, Macro-F1=0.0000 Confusion Matrix: [[0 0 0 60 0] [0 0 0 0 0] [0 0 0 0 0] [0 0 0 0 0] [0 0 0 0 0]]	**Patch 3_0 confusion matrix:** **East Northern Changqing Region** Accuracy=0.0000, Macro-F1=0.0000 Confusion Matrix: [[0 42 0 0 0] [0 0 0 0 0] [0 0 0 0 0] [0 0 0 0 0] [0 0 0 0 0]]

Table 7: *Confusion Matrices for the Supervised LSTM classifier for Patch 0_0 and Patch 3_0*

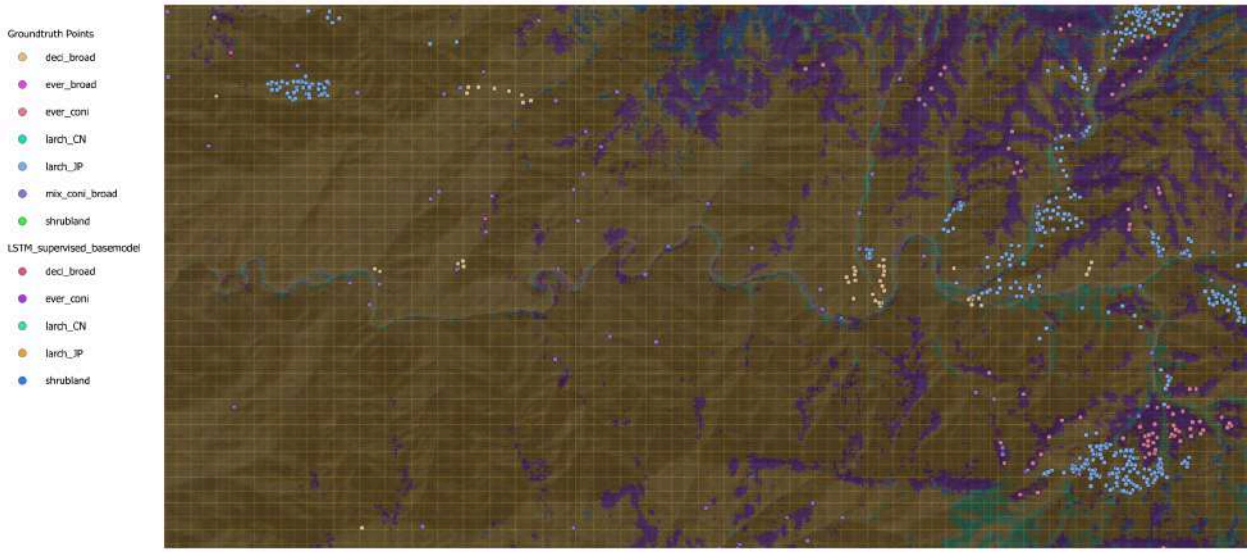

Figure 4: Supervised LSTM Model Overlayed with Groundtruth Points (West Northern Changqing)

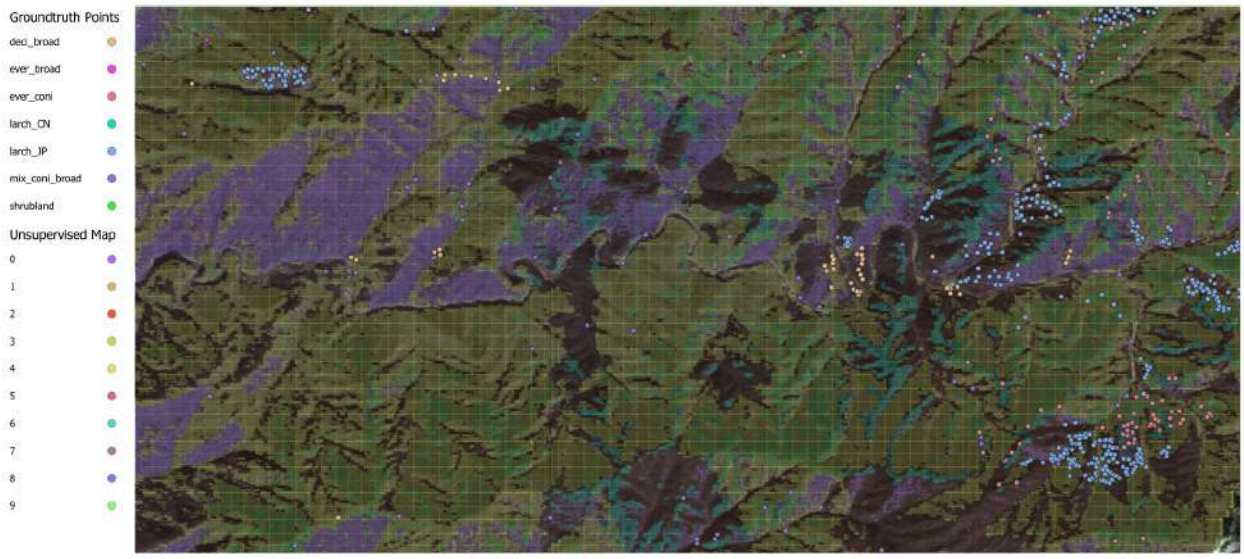

Figure 5: Unsupervised LSTM Model Overlayed with Groundtruth Points in the Same Region as figure 4 (west Northern Chanqging)

The argument here is that while the supervised model excels in regions well supported by training data, it struggles in areas that exhibit locality changes—i.e., regional differences that shift the baseline of the time-series features. In Figure 4, for instance, the supervised LSTM successfully differentiates evergreen conifers (bottom right), a feat also achieved by the unsupervised pipeline. However, the supervised approach simultaneously labels an entire swath of territory as Japanese larch, largely because those “blue” points (both in the bottom right and upper left) are known to be larch_JP. With no alternative groundtruth classes present to contradict it, the model has no incentive to segment other vegetation types. This inflates local accuracy metrics (often approaching 100%) because the confusion matrices in this region show only larch_JP—there are no “competing” classes to reveal classification errors.

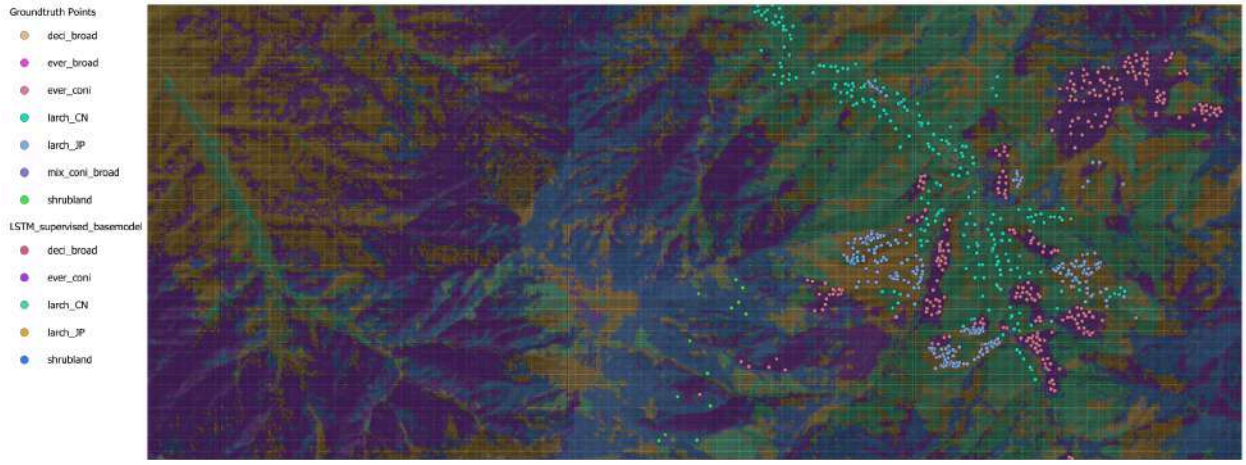

Figure 6: Supervised model performance in areas with good groundtruth (central Huangbaiyuan NNR)

Figure 6 illustrates a contrast: in the central Huangbaiyuan region, where groundtruth data are abundant and diverse (covering ever_broad, larch_CN, larch_JP, etc.), the supervised LSTM performs exceptionally well. This suggests that the 94% accuracy figure holds mainly for regions sharing that same rich data distribution. By attempting to extrapolate the learned time-series patterns from central Huangbaiyuan to more remote areas (e.g., northern Changqing), the model overestimates larch_JP because no equally comprehensive groundtruth set exists there. Consequently, misclassifications surge, underscoring the limited generalizability of a single-region time-series profile. In short, while the supervised LSTM reliably distinguishes classes in well-sampled localities, its predictive power weakens where phenology or terrain differs, and groundtruth coverage is sparse.

Despite these occasional zero-accuracy patches, the overall metrics confirm that the LSTM classifier does learn robust temporal cues—capturing larch_JP's rapid needle flush (spring), peak canopy (summer), and senescence (late fall) distinct from evergreen conifers' stable year-round reflectance. In Huangbaiyuan, where labeled data are extensive, the network indeed aligns well with observed vegetation patterns, achieving near-perfect classification in multiple patches.

4. Extended Discussion and Future Directions

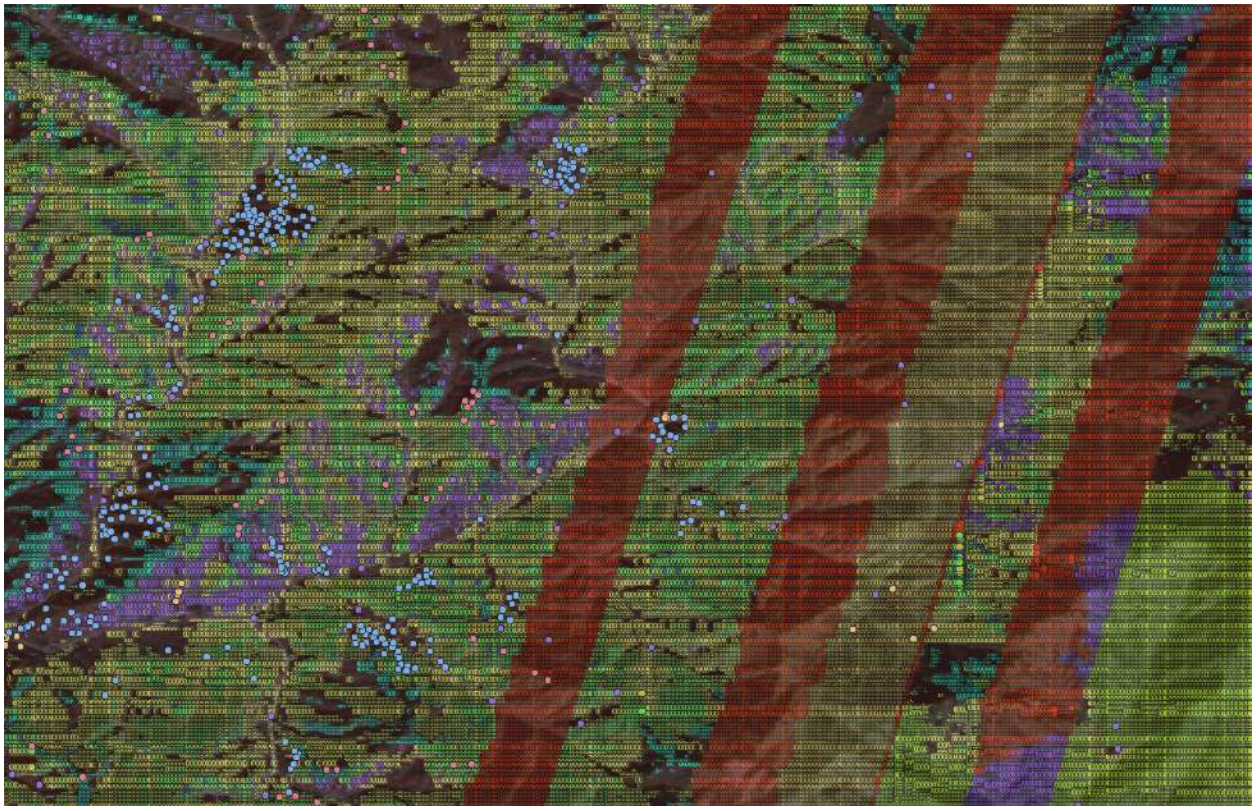

Figure 7: Diagonal “Breaks” in Imagery as Satellite Passes Through Diagonally

A key challenge in assembling monthly time-series data—especially in mountainous regions—is the presence of satellite image gaps as seen in figure 7. Thankfully, this was only present in a minimal region within the 28 patches. The unsupervised model clustered these regions as clusters 2 and 3, which had to be omitted; the same breaks were found in the supervised model as well. It can be deduced that this was most definitely a problem with GEE itself outputting faulty data. Other factors such as cloud cover, shadowing, or scene unavailability introduce breaks in the record too, potentially skewing the reflectance curves used for classification. One promising mitigation strategy is Radial Basis Function (RBF) interpolation, which can reconstruct partially missing time steps by fitting smooth functions to existing data points. RBFs effectively “fill in” data gaps, thereby creating more continuous seasonal trajectories, which Huang et al. (2023) did in their study. Future work may explore systematically applying RBF interpolation to reduce abrupt data drops, detect diagonal “breaks” within the image, and refine monthly phenology signals in larch-dominated or mixed stands.

This leads to the question of what the “perfect scenario” looks like for multi-temporal forest classification. The answer would be having both rich groundtruth inventories with more uniform or moderately sloped terrain. In such settings, advanced deep learning models—like the Transformer architecture used by Huang et al. (2023)—can cope with high-dimensional inputs more gracefully, thanks to large, well-distributed training samples. Transformers also excel at capturing long-range dependencies in time-series, alleviating some of the “curse of dimensionality” that plagues smaller datasets with dozens of monthly features. By contrast, our study faced limited labeled data and rugged terrain in the Changqing region, prompting an LSTM-based approach that, while conceptually reminiscent of Transformers’ sequential encoding, was tailored to a more data-sparse environment. The trade-off is that the model may not generalize to remote localities or unusual phenological regimes as robustly as it would with more comprehensive coverage, as seen in the supervised approach.

Nevertheless, our classification and segmentation results hold practical value for ongoing ecological monitoring. For instance, Professor Binbin Li’s research group conducted extensive field surveys in 2022 & 2023, mapping species presence and stand conditions in Changqing. Zhang (2023), who is also part of the group, and his work reveals patterns of historical deforestation, and the segmentation from this study can further guide the placement of field trap cameras in areas of larch plantation transition. By identifying sharply contrasting zones—such as newly emerged conifer stands adjacent to mixed broadleaf patches—researchers can deploy monitoring devices more effectively to track fauna that prefer edge habitats or forage in young plantations. Ultimately, improved time-series interpolation, multi-year expansions of the dataset, and domain adaptation strategies to handle locality differences will help unify the strengths of LSTM or Transformer frameworks with real-world constraints in biodiversity hotspots.

5. References

AlZuhair, M. S., Ben Ismail, M. M., & Bchir, O. (2023). Soft semi-supervised deep learning-based clustering. *Applied Sciences, 13*(17), 9673.

https://doi.org/10.3390/app13179673

Bremer, L. L., & Farley, K. A. (2010). Does plantation forestry restore biodiversity or create green deserts? A synthesis of the effects of land-use transitions on plant species richness. *Biodiversity and Conservation, 19*(14), 3893–3915.

FAO. (2020). *Global Forest Resources Assessment 2020*. Food and Agriculture Organization of the United Nations.

Fischer, J., & Lindenmayer, D. B. (2007). Landscape modification and habitat fragmentation: A synthesis. *Global Ecology and Biogeography, 16*(3), 265–280.

Goetz, S. J., Baccini, A., Laporte, N. T., et al. (2009). Mapping and monitoring carbon stocks with satellite observations: A comparison of methods. *Carbon Balance and Management, 4*(1), 2.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

Huang, Z., Zhong, L., Zhao, F., Wu, J., Tang, H., Lv, Z., Xu, B., Zhou, L., Sun, R., & Meng, R. (2023). A spectral-temporal constrained deep learning method for tree species mapping of plantation forests using time series Sentinel-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 204*, 397–420. https://doi.org/10.1016/j.isprsjprs.2023.09.009

Hull, V., Roloff, G., Zhang, J., et al. (2014). A synthesis of giant panda habitat selection. *Ursus, 25*(2), 148–162.

Immitzer, M., Atzberger, C., & Koukal, T. (2016). First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing, 8*(3), 166. https://doi.org/10.3390/rs8030166

Jones, H. G., & Vaughan, R. A. (2010). *Remote sensing of vegetation: Principles, techniques, and applications*. Oxford University Press.

Li, Y., Wang, S., Li, X., & Guo, Z. (2020). Habitat selection of giant pandas and its implications for conservation. *Journal of Forestry Research, 31*(6), 2277–2285.

Liao, C., Luo, Y., Fang, C., & Li, B. (2018). The effects of plantation practice on soil properties based on the comparison between natural and planted forests: A meta-analysis. *Global Ecology and Biogeography, 21*(3), 318–327.

Lindenmayer, D. B., & Hobbs, R. J. (2004). Fauna conservation in Australian plantation forests – a review. *Biological Conservation, 119*(2), 151–168.

Loucks, C. J., Lü, Z., Dinerstein, E., Wang, H., & Wang, D. (2001). The giant pandas of the Qinling Mountains, China: A case study in designing conservation landscapes for elevational migrants. *Conservation Biology, 15*(4), 1095–1103.

Maus, V., Câmara, G., Cartaxo, R., Sanchez, A., Ramos, F. M., & de Queiroz, G. R. (2016). A time-weighted dynamic time warping method for land-use and land-cover mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9*(8), 3729–3739.

Newbold, T., Hudson, L. N., Arnell, A. P., et al. (2015). Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science, 353*(6296), 288–291.

Persson, M., Lindberg, E., & Reese, H. (2018). Tree species classification with multi-temporal Sentinel-2 data. *Remote Sensing, 10*(11), 1794. https://doi.org/10.3390/rs10111794

Planet Team. (2017). *Planet Application Program Interface: In Space for Life on Earth*. San Francisco, CA. Retrieved from https://www.planet.com/

State Forestry Administration of China. (2017). *Forest Law of the People's Republic of China*.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment, 8*(2), 127–150.

https://doi.org/10.1016/0034-4257(79)90013-0

Tuanmu, M.-N., Viña, A., Winkler, J. A., et al. (2013). Climate-change impacts on understorey bamboo species and giant pandas in China's Qinling Mountains. *Nature Climate Change, 3*(3), 249–253.

Wang, M., Zheng, Y., Huang, C., Meng, R., Pang, Y., Jia, W., Zhou, J., Fang, L., & Zhao, F. (2022). Assessing Landsat-8 and Sentinel-2 spectral-temporal features for mapping tree species of northern plantation forests in Heilongjiang Province, China. *Forest Ecosystems, 9*, 100032. https://doi.org/10.1016/j.fecs.2022.100032

Wang, Y., Zhang, Y., Song, W., & Zhao, J. (2022). Effects of landscape pattern on biodiversity conservation in plantation forests: A review. *Forest Ecology and Management, 504*, 119792.

Zhang, J. (2023). *Classification of Larch Plantation and Detection of Historical Deforestation Based on Multi-Temporal Interpretation in Southern Qinling Region, China*. Master's Project, Duke University.

Zhang, J., Huang, Y., Pu, R., & Gonzalez-Moreno, P. (2019). Using spectrotemporal indices to improve the fruit-tree crop classification accuracy. *Remote Sensing, 11*(5), 546. https://doi.org/10.3390/rs11050546

6. Signature Work Narrative

Links to thematic courses in DKU: STATS 302 Principles of Machine Learning. STATS 401 Data Acquisition and Visualization. COMPSCI 301 Algorithms and Databases

I first became interested in analyzing larch plantations in the Changqing region during the summer of 2024, when I encountered PlanetScope satellite imagery. At the time, my understanding of remote sensing revolved around single-date approaches and simple band extraction. The possibility of accessing 3 m resolution images from PlanetScope seemed transformative, so I built a pipeline to query, cluster, and download TIF data using their API. This process, which I published to GitHub,

showcased how crucial it is to design robust data workflows rather than merely clicking “download.” Thanks to the lessons from COMPSCI 301 (Algorithms and Databases), I already had a sense of algorithmic thinking—how to structure payloads, manage files systematically, and handle JSON requests to an external service. I also leaned on my experiences from STATS 401 (Data Acquisition and Visualization) to ensure that once I had imagery and metadata, I could produce visual summaries (like quick bar plots of how many scenes I had per month) to verify the pipeline’s correctness.

However, as I dug deeper into the scientific literature, I realized PlanetScope lacked a robust time-series infrastructure and offered fewer spectral bands than Sentinel-2. Meanwhile, Google Earth Engine (GEE) supported monthly composites, cloud masking, and large-scale CSV exports. I also learned that PlanetScope, though providing higher spatial resolution, could not easily deliver the wide red-edge or SWIR bands that my mentors and I needed for phenological analysis. When I saw how Huang et al. (2023) and Wang et al. (2022) used multi-month data to classify forest species, I realized that single-date classification could overlook key changes—especially for larch trees, which lose needles in late autumn and look very different in spring. That discovery shaped the pivot from PlanetScope TIF downloads to building a new pipeline around Sentinel-2 time-series data in GEE.

Because I had already written code to talk to external APIs, the concept of building a multi-step pipeline was familiar. GEE, though, presented a unique challenge: its environment is partially JavaScript-based, but it can also interface with Python. I ended up writing scripts that defined bounding regions, band sets, monthly intervals, and then exported thousands of lat/lon reflectance rows to my Google Drive. From there, I would use local scripts to parse and clean the data—situations where I again applied knowledge from COMPSCI 301 about streamlining data ingestion and from STATS 401 for validating the outputs with visual checks (e.g., plotting NDVI distributions each month to see if they made sense). This entire pipeline gave me a more sophisticated approach to remote sensing: instead of dealing with single images, I had an entire “time-series supply chain” for each pixel over 12 months.

As soon as I saw references in the papers to Recurrent Neural Networks (RNNs) and particularly Long Short-Term Memory (LSTM) networks, I recalled STATS 302 (Principles of Machine Learning) and how it introduced me to basic classification approaches, including random forests, SVMs, and even RNN concepts. I recognized that a random forest might not fully grasp sequential patterns—like how larch's NDVI rises sharply in April and drops in November—so LSTM was perfect for time-dependent data. Building the actual model in PyTorch was daunting: I had to set up CUDA (GPU) drivers, handle large CSVs, and keep my code from crashing. But the fundamentals of vectorization, hyperparameters, and backpropagation were grounded in the theoretical coverage from STATS 302. At first, my model heavily over-predicted the "larch_JP" class, so I introduced class-balanced weights in the loss function. That mitigated—but did not eliminate—overestimation issues.

During this period, I attended weekly three-hour meetings with Professor Binbin's research group, which was largely composed of environmental scientists, IMEP (International Masters of Environmental Policy) students, and a few data enthusiasts like me. These sessions encouraged me to keep iterating on my pipeline. I no longer wanted to procrastinate because each week I would see how others had progressed in R or QGIS, using simpler methods to produce meaningful ecological analyses. While many of them encountered difficulties or slow processing in R, we all recognized that these research questions linked environmental observations (like stand composition, deforestation events, or biodiversity shifts) with data-driven insights—exactly the synergy that a machine learning approach could provide. Seeing how environmental scientists approach field data and how data scientists handle large-scale classification made me appreciate the interplay between design and engineering: we can be like civil engineers and architects working on a complex building project, each side bringing complementary expertise.

In the end, I gathered a labeled dataset of about 4,631 groundtruth samples across seven classes (various larch species, evergreen conifers, broadleaf, and mixed stands). We soon discovered how imbalanced our dataset was: in remote localities,

we hardly had enough lat/lon points for certain classes. As a result, the LSTM model performed far better in the heavily sampled Huangbaiyuan area than in the northern Changqing patches. It was also around this time that I decided to experiment with unsupervised approaches. I read that an LSTM autoencoder could compress the time-series into latent vectors, then cluster them with a Bayesian Gaussian Mixture Model. This method let me see whether some conifer stands in under-sampled localities might differ from the “larch_JP” that my supervised model insisted upon. Indeed, the unsupervised segmentation discovered multiple distinct spectral-temporal signatures where the supervised model saw only one.

Another intricacy I encountered was the concept of “locality shifts.” Sometimes, 60 km away from Huangbaiyuan, the baseline reflectance or NDVI would differ enough that the supervised model guessed “larch_JP” across an entire area. By exploring the autoencoder clusters, I saw how the local climate or microtopography likely altered the reflectance curve—something our small groundtruth set could not adequately represent. In the future, we could adopt domain adaptation or multi-year composites to reduce these shifts, or apply Radial Basis Function (RBF) interpolation to fill data gaps more consistently.

Reflecting on the entire journey, I can see how my STATS 302 introduction to RNNs, COMPSCI 301 training in algorithmic thinking and data structures, and STATS 401 emphasis on data visualization collectively shaped this Signature Work. Stats 302 gave me the theoretical bedrock to appreciate sequential models. Compsci 301 ingrained robust pipeline design—whether hooking into the PlanetScope API or GEE’s CSV exports. And Stats 401 reminded me that data must be communicated visually for trust and clarity: I continually used bar charts, time-series plots, and confusion matrices to assess the data’s shape and the model’s performance.

Meanwhile, these weekly research group meetings gave me not just accountability,

but also a genuine appreciation for the environment. Listening to IMEP students talk about habitat corridors, biodiversity surveys, or how certain understory vegetation might flourish in new plantations made the concept of “just a classification map” feel far more alive. Understanding that my results could help them decide where to deploy camera traps or where to focus reforestation was highly motivating. It reinforced the fact that data science is not an isolated discipline—it thrives when integrated with domain expertise.

Ultimately, the final classification pipeline for the Changqing region produced robust results in well-sampled areas, yet highlighted overestimation in remote patches. Such nuance underscores the interplay between groundtruth coverage and model generalization. My shift from PlanetScope to Sentinel-2 shows how carefully one must weigh different satellites’ spectral and temporal capabilities. And the design of LSTM pipelines—plus the addition of an unsupervised clustering branch—demonstrates the creative methods one can adopt to tackle data scarcity. The project’s outcome, a time-series classification map, not only advanced my personal skill set but also provided environmental policy students and ecologists with an additional lens to plan field deployments. Sharing these results at the weekly meetings was a deeply rewarding experience, merging my data-centric perspective with their ecological knowledge.

In conclusion, this SW project taught me the value of bridging advanced algorithmic methods (LSTM, autoencoders) with an appreciation for real-world applications (protecting larch stands, guiding biodiversity initiatives). The synergy of multiple thematic courses—machine learning theory, data engineering fundamentals, and data visualization—made the pipeline feasible. And above all, interacting regularly with environmental scientists opened my eyes to the bigger picture: that time-series remote sensing is more than a coding exercise; it is a collaborative tool for understanding and managing dynamic, ecologically vital landscapes.