Modeling Forest Transition Using Deep Learning: A Remote Sensing Study of Taucamarca, Peru

Sadie F. Calhoun and Clara R. Atwell

Dr. G. Andrew Fricker

A senior project submitted to the Department of Social Sciences, Cal Poly, San Luis Obispo

in Partial fulfillment of the requirements for the degree of

B.S. in Anthropology and Geography

March 2023

Author Note

Special thanks to Cal Poly Latin American Studies program for funding this research and to Dr. Jim Keese and G. Andrew Fricker for all your support on this project.

Abstract

A history of intensive lumber use in the Peruvian Andes has led to wide-spread deforestation. Wood is a vital resource to the Indigenous people of the rural Peruvian Andes. In primarily subsistence rural economies, it is depended on to supplement incomes and is used locally in construction, heating, and cooking. Since 2015, previous qualitative research has investigated various components of the political ecology of eucalyptus, pine, and native Andean species in Taucamarca, Peru. The purpose of this research is to quantify tree species land-cover and maturity in Taucamarca from high-resolution satellite imagery. We used the deep learning frameworks integrated in ArcGIS Pro to classify land-cover. The model was trained with hand-digitized samples from satellite imagery obtained in 2018 and 2022, with assistance from on-the-ground GPS points that were collected in August 2022. Since this project focuses on a large-scale pine reforestation project in Taucamarca, we aim to allow for monitoring future land-cover changes in the region.

Introduction

Peru is a middle-income country with a high degree of wealth inequality. Poverty within the country is concentrated between the urban poor and rural indigenous populations. Twenty-six percent of Peruvians are Indigenous and depend on a primarily subsistence economy based on the cultivation of crops such as potatoes, corn, wheat, barley, quinoa, and vegetables, along with small-scale livestock production (Keese et al., 2020) Since the 1960s, it has become increasingly common for local incomes to be supplemented by joining small-scale logging operations (Keese et al., 2017)

Wood is a vital resource to people in the Peruvian Andes. Its production is depended on for income, heating, and 75% of cooking needs (Keese et al., 2017). This heavy use, coupled with other factors such as mining and monocropping discussed later in this paper, has led to the deforestation of 90% of the native Peruvian Sierra Click or tap here to enter text. Since the 1960s, blue gum eucalyptus (*E. globulus*), and more recently two Mexican pine species (*Pinus radiata and Pinus patula*), have been promoted by international organizations and the Peruvian government to reforest large tracts of land and meet wood consumption needs (Luzar, 2007; Raboin & Posner, 2012).

In 2015, 2018, and 2019 Keese conducted field research on cookstove usage in five rural Peruvian towns, including the study area of this project, Taucamarca. Harvesting of wood for cooking and heating has been a significant historical source of deforestation. From this research, Keese concluded that the planting and harvesting of trees plays a central role in the dynamic of social and economic life in the Peruvian Andes (Keese et al., 2020). In 2022, Keese took a sabbatical to further investigate the political ecology of pine and eucalyptus in the Peruvian Andes. His research builds off Luzar's 2007 research *The Political Ecology of a "Forest*

Transition": Eucalyptus forestry in the Southern Peruvian Andes. Both Keese and Luzar noted in their work that there is a need for future research that quantifies the extent of the pine and eucalyptus forest transition in the Peruvian Andes. They argue this work is a key component to understanding the farming dynamics of communities in the Peruvian Andes.

The goal of this study is to meet this research need. We aim to "provide insight into the spatial aspects of Peru's forest transition," (2007). While Luzar and Keese's work seeks to explain the layered social, political, and economic reasons why pine and eucalyptus are being planted at such large scales in the Peruvian Andes, our work seeks quantify pine versus eucalyptus cover in Taucamarca Peru using ArcGIS Pro's integrated deep learning package.

Since 2019, students have been working on the remote sensing side of this project. However, they have never used deep learning to assist in distinguishing between pine and eucalyptus. Only one of three previous projects involved collecting ground truthing points, which is critical to understanding imagery in small areas of interest, and this student's work failed to produce readable data. As a result of these factors, a central goal of our research is to create readable data that can easily be reproduced and built upon in future student projects.

Area of Interest

Taucamarca is a 22 square-kilometer indigenous Quechua speaking community in the Cusco region (the equivalent of a U.S. state), located roughly 45 kilometers east of Cusco. According to community members interviewed in August of 2022, the community has a population of 320 (74 families). The base of the community sits within a valley at 3,700 meters. Its highest points reach roughly 4,400 meters in altitude, 18 meters (60 feet) below the summit of Mt. Whitney, the highest point in the continental United States. Because the Andes lie within the tropics, the climate of Taucamarca is dry and temperate, making it suitable for growing crops, raising livestock, and more recently forestry.

Taucamarca was chosen as the area if interest for this study for two reasons, Cal Poly's connection to the locality through Keese's LPG cookstove research and the town's large-scale pine planting program. This program is funded by the municipality of Caicay in which Taucamarca resides. It began in 1999 when a few small pine plantations were planted as an alternative to Eucalyptus which has been used since 1960s for local use and to supplement local

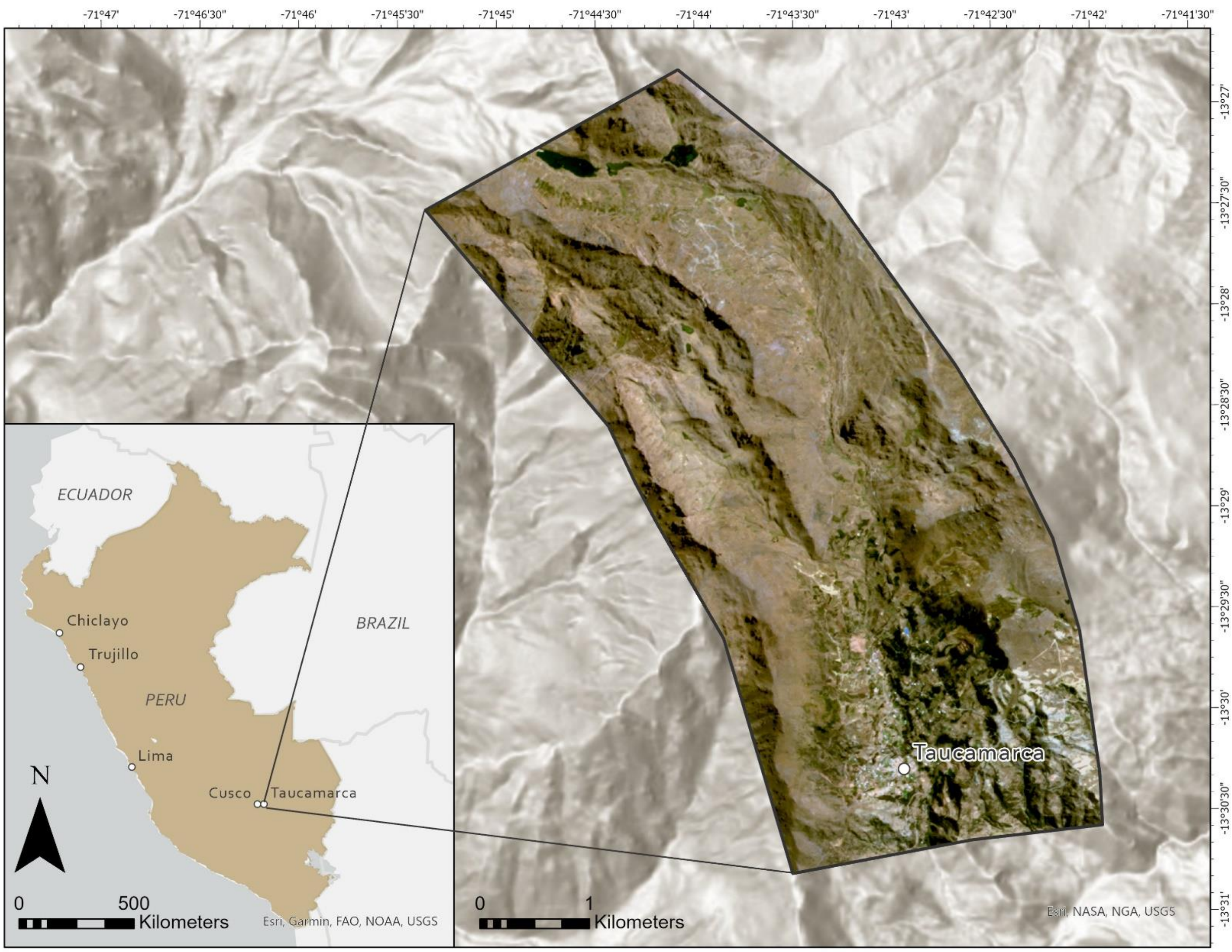


Figure 1 *Map of area of interest, Taucamarca, Peru. Taucamarca is a 22 sq km remote Andean town located 45 kilometers east of Cusco. Image source: Planet Labs 08/17/2022.*

incomes. Since 1999, three more large-scale plantings have taken place from 2012 to 2013, in 2015, and most recently in November 2023. According to local forestry workers in August of 2022, over the next three years, 120,000 more pine plants will be planted in the community. Given the small population and physical size of this community, this is an extremely large-scale project. Today, pine dominates the landscape of Taucamarca, especially in the upper reaches of the community where other plants growth is stunted.

Study Context

The Peruvian Andes are part of the Tropical Andes biodiversity hotspot which is home to 12-17% of the world's species (Keese et al., 2007). Centuries of human dependence upon this fragile landscape has led to its near complete deforestation. Since the 1960s, government and international organizations have promoted reforestation of the Andean Sierra with non-native Eucalyptus *(E. globulus*) and later two native Mexican pine species *(Pinus radiata and Pinus patula)* (Raboin, 2012). The roots of Peruvian deforestation, historically blamed on Indigenous small-holder agriculturalists (Ravikumar et al., 2017), can be understood through a political-ecology framework, which views environmental changes in the context of larger political and economic processes (Luzar, 2007). The political ecology framework coupled with the forest transition model can then be used to explain afforestation in the Peruvian Andes. In this paper, these frameworks are used to contextualize this remote sensing study quantifying tree speciation in Taucamarca, Peru.

Forest Transition in the Peruvian Andes

Since colonization, deforestation has been the dominant forestry trend in the Peruvian Sierra. The main drivers of forest loss are clearing for agriculture and pasture, gold mining, timber, and demand for firewood (Keese et al., 2007; Tito et al., 2022). Since the 1960s, large-scale reforestation efforts promoted by the state and international development organizations have greatly impacted the way in which individual farmers and communities use their land. This process of afforestation after long periods of forest decline has become known as forest transition and tends to follow a distinct trend captured within the forest transition theory (Mather & Needle, 1998).

Forest transition theory predicts that urbanization creates a growing consumer base in cities, which leads to an increase in demand for wood resources. This urbanization simultaneously causes a decrease in rural agrarian labor, which pushes farmers to transition to less productive areas of their land to low-intensity tree farms (Mather & Needle, 1998; Rudel, 2009). In the case of Andean Peru, where crops cannot grow at high altitudes, this also involves tree plantations replacing a tropical alpine grassland system known as jalca (Tovar et al., 2013b). This process is reinforced by government and NGO programs (Rudel, 2009),

Remote sensing has proved to be a powerful tool for measuring land use and land cover change (LUCC) and forest transition in the Andes (Keese et al., 2007; Tovar et al., 2013). However, it is difficult to draw definitive conclusions about the causes of LUCC without the collection of on the ground data and a complex understanding of the political ecology of an area (Ravikumar et al., 2017). For this reason, this research solely seeks to measure the extent of Pine and Eucalyptus plantations. Keese is currently working on a paper that explores the political ecology of Taucamarca's forest transition in depth. We spent the month of August 2022 in the study area following Keese in his interviews with local forestry workers, Peruvian forestry NGOs, local and provincial government officials, brick makers in Cusco, and lumber yards in Cusco. We additionally spent time ground truthing in Taucamarca and gaining an in-depth understanding of the local land use and cover. This time spent on the ground in Peru was essential to understanding the significance pine and eucalyptus in social and economic life in the Andes.

History of eucalyptus and pine in the Peruvian Andes

Blue gum eucalyptus (*E. globulus)* was first encountered by Europeans in Australia in 1770 (Dickson, 1969). The tree was quickly spread around the world to tropical, subtropical, and Mediterranean climates. In the late 1800s, it reached Peru where it was initially grown at estates or in gardens (Dickson, 1969).

In its native Tasmania, eucalyptus's altitudinal limit for growth is 400 meters. However, near the equator it can grow rapidly at altitudes between 2,000 and 4,000 meters without irrigation (Dickinson, 1969). This makes the eucalyptus and ideal species for forestry in the Andes where native trees grow slowly and with twisted wood. The species additionally meets lumber and firewood needs that the native trees cannot. Eucalyptus has the ability to coppice, meaning that the trunk rapidly produces new cuttings after being cut down. It also is a hot burning wood. It is used for both cooking and heating in the rural Andes and as a cheap fuel source for firing roof tiles and bricks (Luzar, 2007).

While eucalyptus is undeniably a practical source of fuel and timber, it is also now taking over native landscapes and leading to environmental degradation (Luzar, 2007). Tree plantations were found by Tovar et al. (2013) to be the second greatest contributor to lost jalca grasslands after overgrazing. These native grasslands are critical to regional hydrological regulation (Raboin & Posner, 2012). The tree's high consumption of water is also believed to reduce local stream flows. Eucalyptus absorbs 10 to 25% of the water that passes through it (Luzar, 2007). It additionally, outcompetes other plants, leads to soil acidification, and the high levels of oils in leaves can leach out in waterways impacting aquatic life (Luzar, 2007).

Eucalyptus planting began to expand at a large scale during the 1960s as the state became increasingly involved in land holdings and reforestation efforts. In 1963, with help from USAID loans, the state sponsored eucalyptus plantations aimed at providing a cheap source of mine supports (Luzar, 2007). The 1969 agrarian reform under the presidency of Juan Velasco Alvaro provided a dramatic shift in Peruvian land ownership and marked the beginning of large-scale government involvement in Eucalyptus forestry. Prior to the agriculture reform, 3.9% of Peruvian farmers controlled 56% of the country's arable land. The reform limited the maximum hacienda holding to just five hectares (Luzar, 2007). During this period, the government provided communities with credits to plant eucalyptus, technical assistance, seedlings, and/or tools for planting.

These reforestation efforts were aimed at meeting the state's demand for firewood and timber needs. However, the high-level of state assistance led rural peasants to enter a patron-client relationship with the state (Luzar, 2007). This relationship was interrupted in the 1990s when structural-adjustment programs led to the large-scale privatization of government-sponsored programs. This period led NGOs to step into the spheres previously occupied by the government, in Peru this included Eucalyptus forestry (Luzar, 2007). Today both the state and NGOs sponsor reforestation initiatives pushing for large-scale planting of eucalyptus, pine, and native trees.

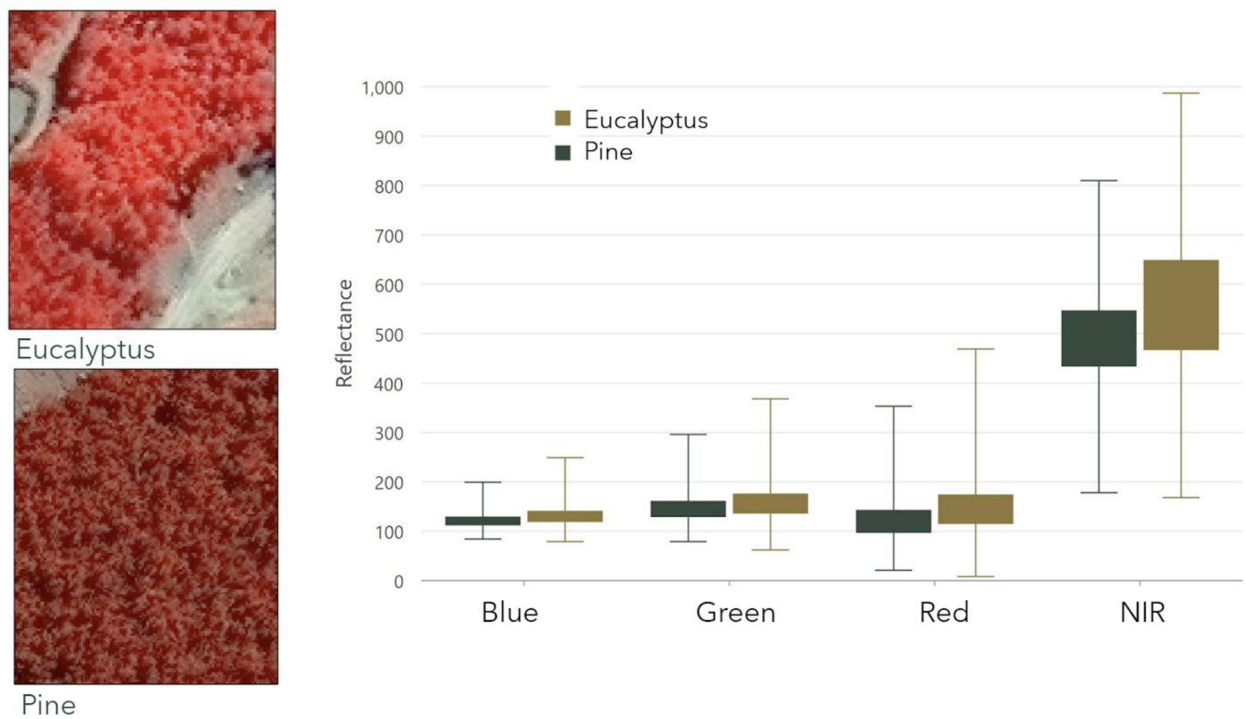
Pine species (*Pinus radiata and Pinus patula*) were first planted in Peru in the 1940s and began to be promoted by the Ministry of Agriculture for plantations in the 1990s (Guariguata et al., 2017). Pine proves to be more adaptable to the marginal lands of the Sierra than eucalyptus. It grows well on slopes at altitudes where eucalyptus would otherwise be stunted. Additionally, it grows to maturity quickly and produces strong timber (Keese, 2022). The strong roots of the trees also prevent erosion (Keese, 2022). Similar to eucalyptus, pine species use significantly more water than native plants leading to reduced available water for irrigation (Raboin et al., 2012). Compared to native jalca grasslands, the species was found to lead to 63% less soil water retention (Tito et al., 2022). Pine plantations also lead to reduced surface runoff (Tito et al., 2022).

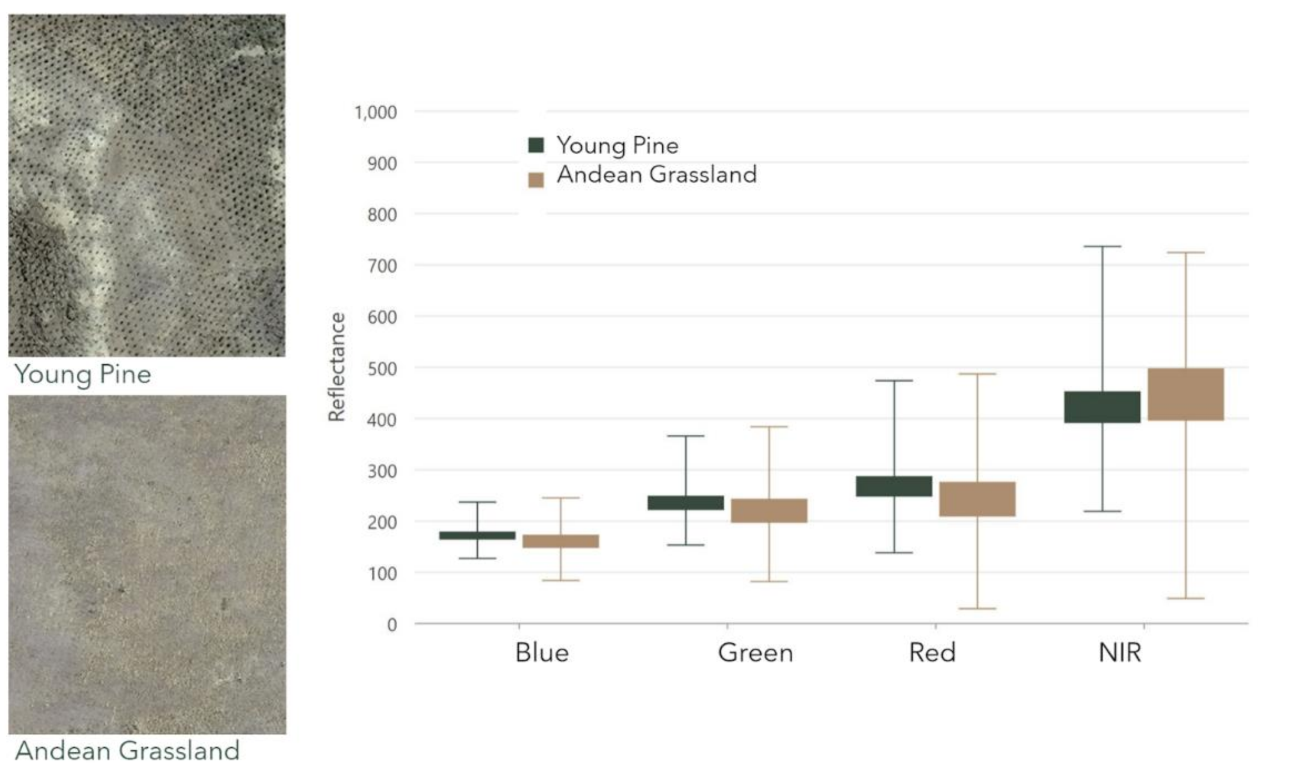
Remote Sensing

Using deep learning on satellite imagery is a relatively new development, with applications beginning only in the past decade. More recently, Esri released its deep learning frameworks, which makes these technologies more accessible in a pre-packaged form. Deep learning takes over from its predecessor “Supervised Classification” in ArcGIS. While supervised classifications are accurate and useful in differentiating between different land uses (Alo et al. 2008), it is only able to look at an individual pixel’s spectral signature or an area’s average. This type of classification is a linear, single layer algorithm. It also only allows for one time training of data. Deep Learning classifications, on the other hand, are able to take into account the area surrounding pixels, meaning it is able to “see” patterns and textures. It is non-linear, creating a multi-layer convolutional neural network that allows for continuous training.

While senior projects before us attempted to use supervised classifications, this proved difficult for the area and species. Intermediate pine and eucalyptus have similar spectral signatures, as well as young pine and Andean grassland (Figure 2). By using a convolutional neural network (CNN) for our attempts to classify trees, we can take into account textures and patterns in the imagery. In the literature, there are many attempts in classifying land cover in satellite imagery with CNNs. The most accurate of these attempts have been to classify plants in a monoculture fashion, such as oil palm and mangroves (Mubin et al 2019; Weinstein et al 2019) which have clear shapes and patterns and do not experience mixing of plants. Also accurate in the literature is the application of CNNs in classifying land cover (Sefrin et al 2021; Kasul et al 2017). These involve broader classes of land, such as urban, grassland, and trees, which have more distinct spectral signatures, making it easier for a CNN to accurately classify. Projects attempting to differentiate between stands of species as similar as pine and eucalyptus are limited to projects utilizing drones, and those with mass amounts of data (Nezami et al. 2020, Bolyn et al. 2022).

This research is novel because nearly all of the existing literature which applies deep learning to satellite imagery uses pre-existing annotated data. These projects are on well documented areas in more economically developed areas, where there is longer standing geospatial data. They are also generally in conjunction with the Parks and Recreation, or other governmental organizations (Hartling et al 2019, Sefrin et al. 2021, Bolyn et al. 2022). Our project, on the other hand, is focusing on a very remote area of the world, where there is little geospatial data or research. The political and ecological implications that motivate this Remote Sensing project also distinguish it. Other projects are also mostly interested in groundbreaking advances in ML, whereas this one is more motivated by the monitoring of forest transition in Taucamarca.


Eucalyptus
Pine
1,000
900
800
700
600
500
400
300
200
100
0
Reflectance
Eucalyptus
Pine
Blue
Green
Red
NIR


Young Pine
Andean Grassland
1,000
900
800
700
600
500
400
300
200
100
0
Reflectance
Young Pine
Andean Grassland
Blue
Green
Red
NIR

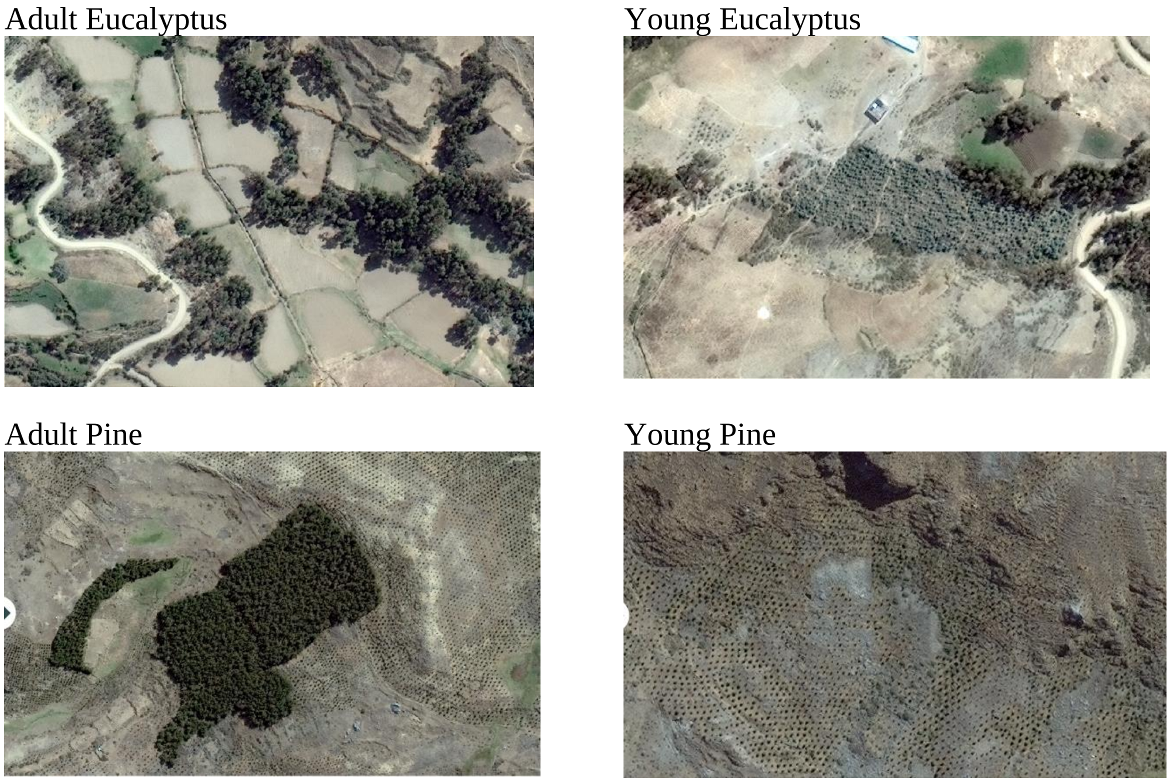


Figure 2a, 2b. and 2c 2a: Intermediate growth Eucalyptus and Pine have very similar spectral signatures. 2b: Young pine signatures are very similar to andean grass land, as they are planted on the grasslands. 2c: Four classes of forest compared.

Methods

Ground Truthing

In August 2022 we traveled to Cusco Peru for a month with Keese on a Cal Poly Latin American Studies grant to collect ground truthing data and define our area of interest. Prior to this trip, we had trained a preliminary model based on Worldview imagery from September 6, 2018. The shapefile for this imagery was a rough guess of community boundaries. Our first priority in Taucamarca was redefining this area of interest according to local's perception of community boundaries. After discussions with local forestry workers, we established the southern base of the community was a "Taucamarca" sign on the road to the community. The northern edge is at the farthest edge of their future pine plantings. The eastern and western community borders are the ridges of the valley Taucamarca lies in. Figure 4 depicts the redefined borders of the community along with our ground truth points.

Data points were collected with two EOS Arrow GPS receivers in the ArcGIS collector app. For each point, we collected data on date, agriculture type, forest type, landcover, and tree growth stage. Additionally, we took detailed comments and photos at each ground truthing point which were saved in the app. Although we initially planned on collecting 150 ground truth points at randomized locations, we quickly realized this was not feasible.

Our ability to collect data at these randomized locations was restricted by altitude, access, and time. At 3,500 to 4,500 meters elevation, our oxygen was restricted and there were rarely paths up the sides of the steep marginal slopes in which trees are planted. These slopes became especially precarious during a few rainstorms experienced in the field. Additionally, some tree plantations were located on farmers' private plots. Our greatest barrier to data collection however was access. Taucamarca is located 45 kilometers outside of Cusco, where we were staying. However, the drive there could take anywhere from an hour to two and a half hours depending on road conditions and traffic. Half of the drive is along narrow dirt roads subject to rockslides that completely block access to the road. On our las trip to the community in late August, we experienced one such rockslide that required an additional hour and a half detour to reach Taucamarca.

The remote nature of the community also required us to have all maps of the community pre-downloaded and organized in the ArcGIS collector app. ArcGIS collector only allows download of area maps in less than one-kilometer by one-kilometer squares. This necessitated the pre-download of over 22 map squares each. Small issues relating to finding the square that corresponded with our GPS location ended up consuming far more time in the field than we initially predicted.

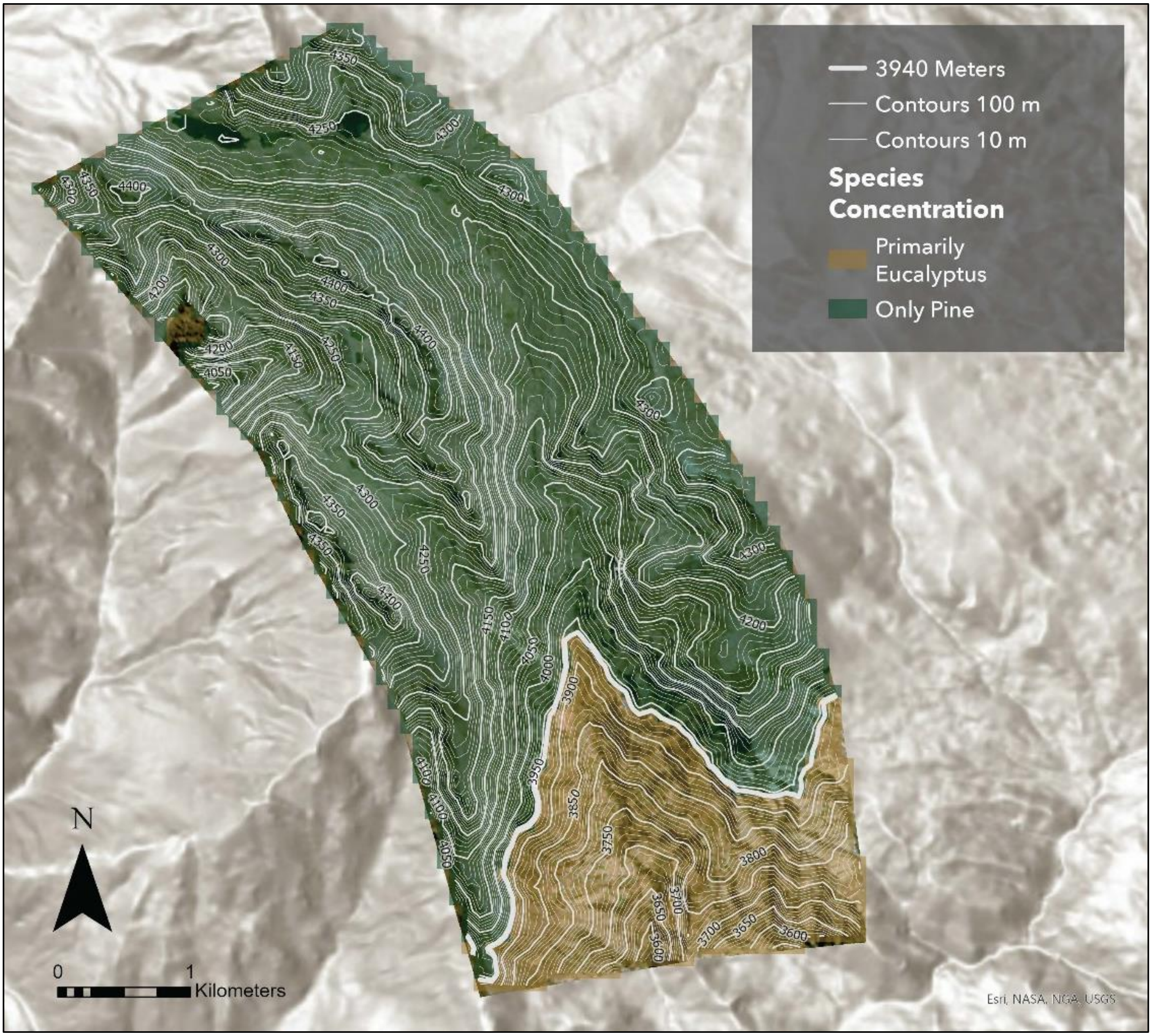


Figure 3 Tree species concentration by elevation. Locals primarily plant eucalyptus in the base of the valley. Pine plantations are concentrated above 3,940 meters where eucalyptus and crops are stunted in growth. We concentrated our ground truthing along this elevation gradient. Data Sources: STRM Void Filled, USGS; Planet Labs (8/17/2022).

Given these restraints, our ground-truthing process became focused on gaining a sense of place in area of interest. Opposed to collecting data in areas with clear concentrations of pine or eucalyptus, we targeted collection in areas of significant transition between the two species where eucalyptus and pine were located directly next two each other. The most significant gradient between the two species was at the elevation line of 3,940 meters. Below this line, eucalyptus dominated the landscape with small patches of pine interspersed on the most marginal tracts of land. Above this line, eucalyptus is stunted in growth and therefore not grown. Our final ground truth data set is made up of 149 points taken primarily along this elevation gradient as seen in Figure 4.

In addition to collecting ground truth data, we flew a DJI Phantom drone at the community base and along the 3,940-meter elevation line. This drone imagery provides higher spatial resolution imagery than our satellite imagery that is useful both in presentations and for acquainting future students to the study area.

Deep Learning Methods

A pixel-based U-Net learning model was used after several trials comparing object and pixel-based classifications. We found object-based learning to be unsuitable after several failed trials, likely due to the variable nature of plantation shape and size. U-Net pixel classifications has also been proven the most successful compared to other CNNs in other projects attempting to differentiate between land cover (Zhang et al. 2020).

The satellite imagery used to train and test the deep learning packages is described in Table 1. These images were selected because they are in the dry season, when there is the strongest contrast between the forest vegetation and the grasslands which are mostly dead and brown. 50 cm was the highest satellite imagery available. Lower resolution, open-source imagery does not have adequate resolution to confidently differentiate between pine and eucalyptus, so it was critical to have this higher resolution imagery.

Table 1

	Sept. 6 2018	July 9 and Oct. 12 2022
Satellite	Worldview-3	Jilin-1
Resolution	50 cm	50 cm
Bands	4-band (RGB+NIR)	4-band (RGB+NIR)

First, a training sample schema was created. This included 6 classes, non-forest, young eucalyptus, intermediate eucalyptus, young pine, intermediate pine, and native (to be used in future projects). This training schema was saved and added to a document to be shared with future students working on the project to be able to continue training the model to improve accuracy.

Next, training samples were created using the *Training Samples Manager* tool. Initially, the entire September 2018 Worldview-3 50 cm resolution, 4 band imagery was annotated for pine and eucalyptus and partially annotated for non-forest using the Peru Agroforestry Schema. These were saved as a feature class and added to the project's map. Next, using the export training data for deep learning, these 2018 annotations were used to create training samples. The tile size and stride were 256 and 128 respectively. These were processed in the TIFF image format and Classified Tiles Metadata Format, since they are for a pixel based classification.

Next, the deep learning model was trained using the deep learning framework within ArcGIS Pro. The U-Net (Pixel classification) tool was used with the default batch size of eight, and default class balancing, mixup, focal loss, and chip sizes. Next, this .dlpk was used to classify the 2022 imagery by running the Classify Pixels Using Deep Learning tool.

After assessing the accuracy of the model trained on the 2018 imagery and run on the 2022 imagery, it was seen that it was over classifying young eucalyptus. To combat this, the model was trained again with samples from the 2022 imagery, which was Jilin-1 50 cm resolution 4-band imagery. This process was slightly different, where newly drawn training samples (with the same schema from before) are fed into a pre-trained model. This is made possible by the Train Deep Learning Model tool in ArcGIS Pro.

In order to calculate the accuracy of the model pre and post training on the 2022 imagery, 100 random points were generated in the map extent. These were labeled for what they were classified as, and what they were from looking at the imagery, and the counts of these were used to calculate a kappa coefficient for two models: the one trained on 2018 imagery and run on 2018 imagery (Peruforest) and the model trained on the 2018 and 2022 imagery (Peruforest2).

Results

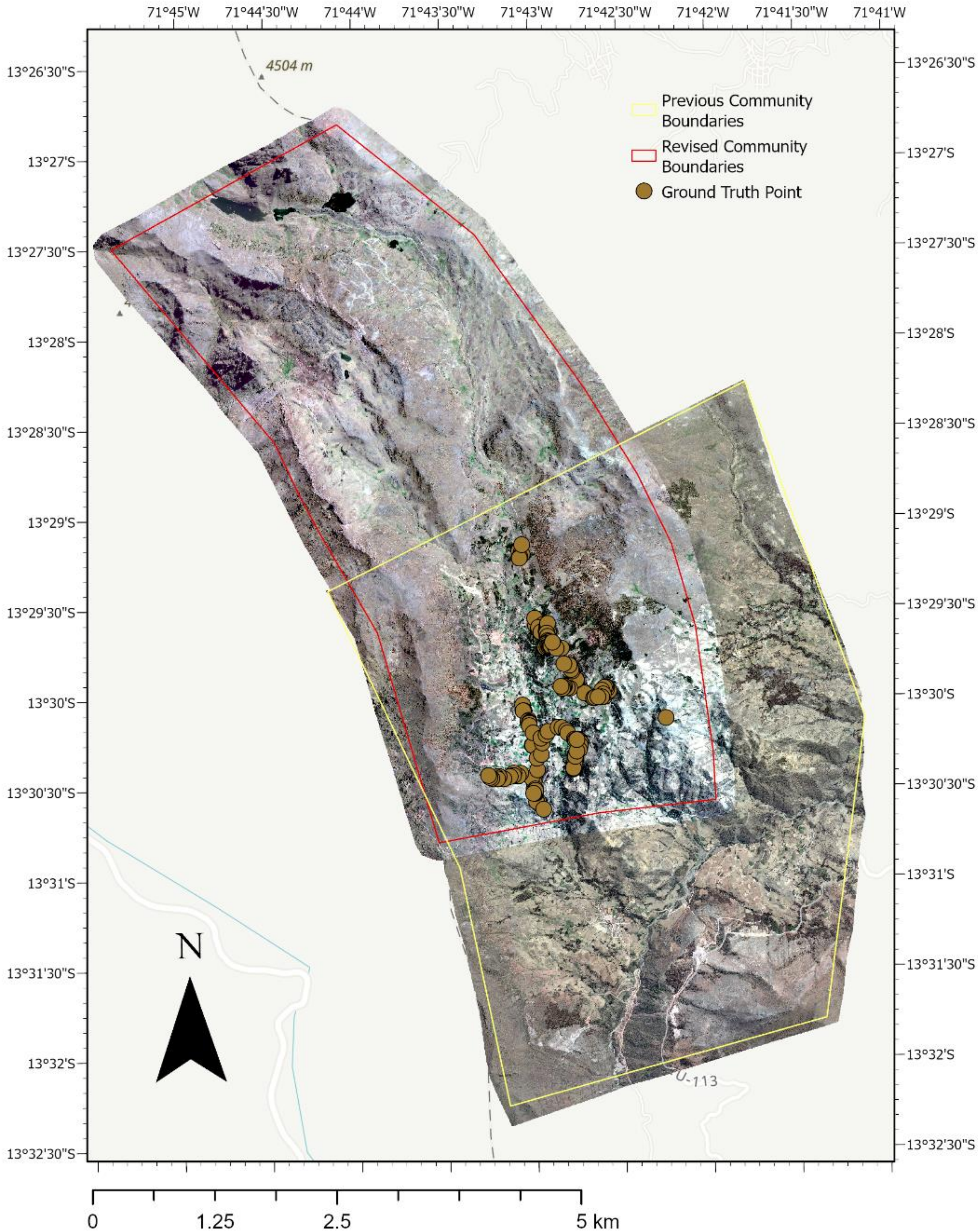


Figure 4 Previous and updated community boundaries along with accompanying satellite imagery. Ground truthing points from August 2022 overlayed.

The first and second models trained gave good overall accuracy, and poor to okay kappa values. First we will assess Peruforest's performance, followed by Peruforest2. Full accuracy assessment tables are available in the appendix.

Peruforest had an overall accuracy of 91% and a kappa of -0.48. This model was very good at classifying young and intermediate stands of pine. There are very large areas of young pine planted, and the model was able to pick up on these large areas. There are smaller patches of intermediate pine (pilot pine plots planted in 1999), and these are fully filled in. The model is very successful at classifying these, including their mostly irregular boundaries. Peruforest, however, was not as accurate at classifying eucalyptus. It is accurate in identifying intermediate eucalyptus, however is not as good at labeling the stands all the way to their boundaries. It tends to cut in from the boundary some. The model also struggles with young eucalyptus. The model tends towards classifying stands of completely young eucalyptus as either a mix of young pine and eucalyptus or as not forest.

Peruforest2 run on the 2022 imagery had an overall accuracy of 84% and kappa of -0.17. This model was very good at classifying pine, both young and intermediate. It picked up on the boundaries of stands of forest and fully classified the insides of those boundaries for the most part. However, it was very poor at finding eucalyptus and pine, as it did not classify a single case of either in the imagery. There are several places where eucalyptus is actually classified as pine, and the rest are not classified (not forest). When Peruforest2 was run on the 2018 imagery, the results were very poor, too poor to make it worth running a kappa score on the classified raster (Appendix 2).

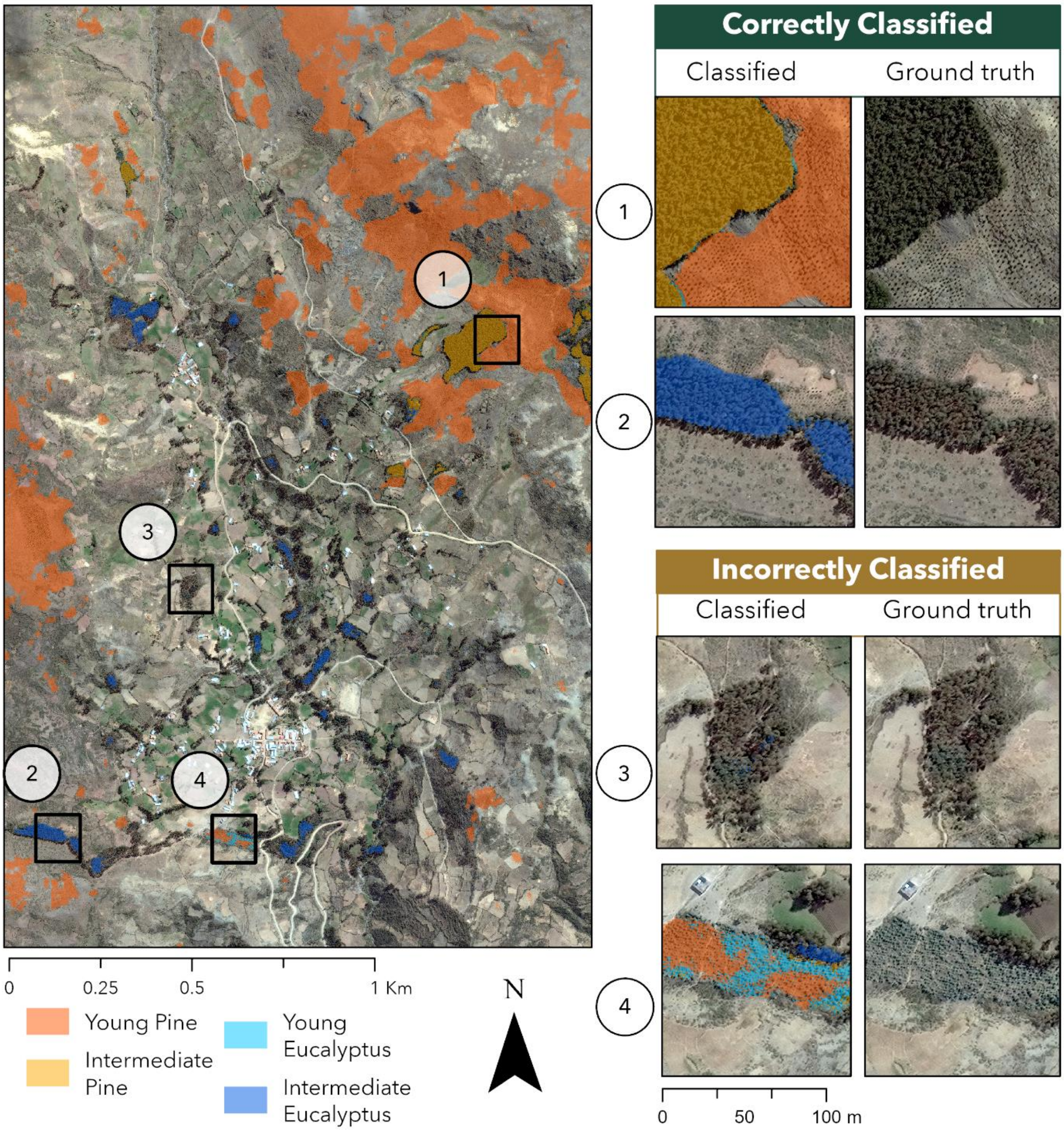


Figure 5 Peruforest model, trained only on 2018 imagery. (Panel 1) A stand of intermediate and young pine correctly classified, the boundaries very accurate. (Panel 2) A stand of eucalyptus accurately classified. (Panel 3) Mixed young and intermediate e

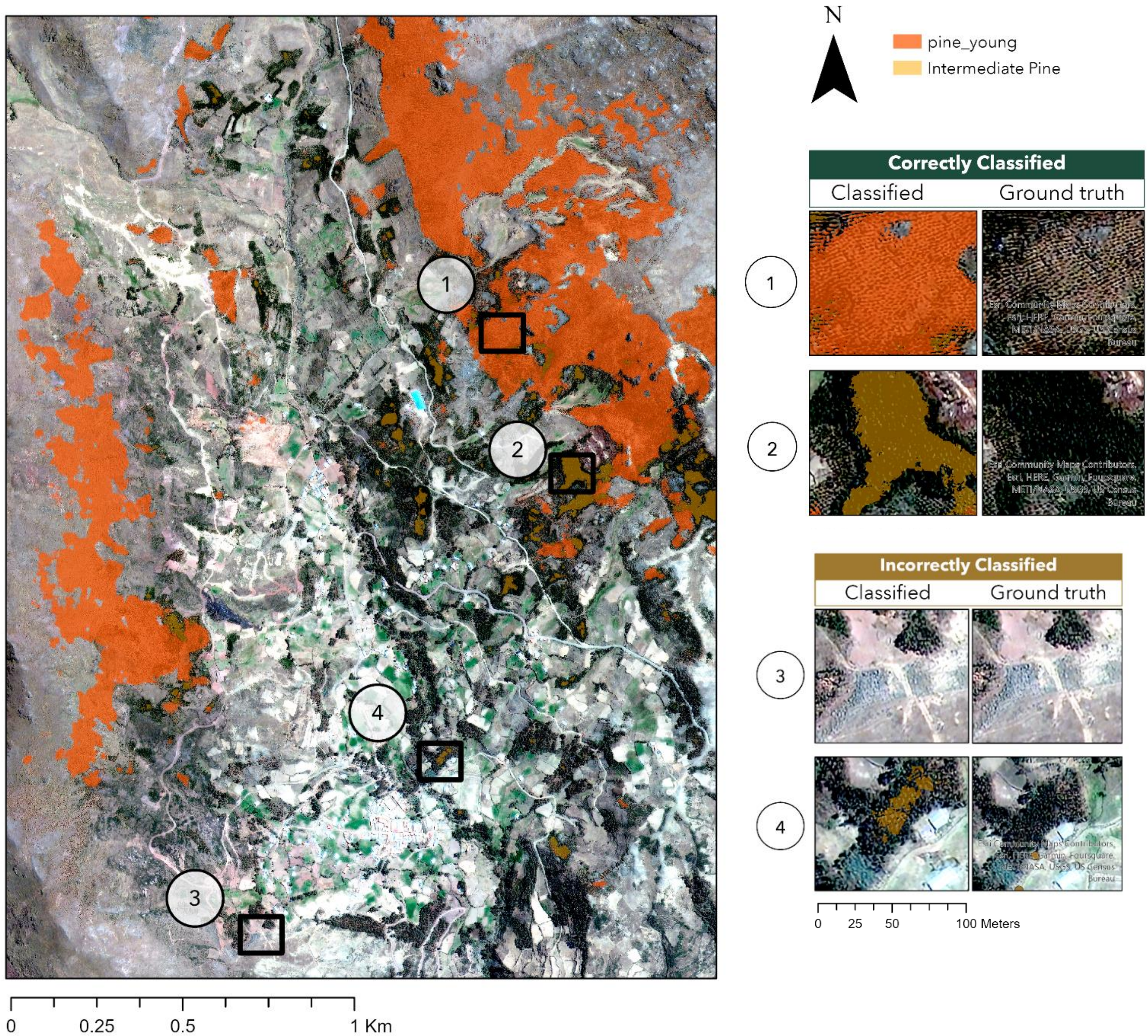


Figure 6 Peruforest 2 model, trained on both the full 2018 and partially on 2022 imagery and run on the 2022 imagery. (Panel 1) a stand of young pine is classified accurately, and successfully does not classify the small patch of no pine within the stand. (Panel 2) A stand of old pine is correctly classified, although the margins of the pine are not fully classified. (Panel 3) Young Eucalyptus is not classified as anything (not forest). (Panel 4) Old eucalyptus classified as old pine.

Discussion

The high overall accuracies and low kappa scores for the two deep learning models can be explained by the method of assessment given the size of the imagery, and a lack of adequate training for all of the variation in tree appearance.

The negative kappa scores are due to the fact that the overall accuracy was lower than the expected accuracy. The lowness of the Kappa scores is partly because the imagery area is so large (25 square kilometers) and there is only tree cover in a small portion of that. Most of the random points generated therefore fell on non-forest, which were mostly currently unclassified (97% and 98% for 2018 and 2022 respectively). This led to a lack of diversity and spread in points - there were less than 10 for any other class while there were 78 and 82 for non-forest. A more accurate assessment of future iterations of the model would likely be to generate a random number of points within each class of the classification and assess each of these.

Visual, qualitative assessments of the classified rasters from the model gives a better idea of how good the model is at classifying imagery and fitting for the shapes and edges of plots of forest. One very promising observation is that the model is not classifying areas of the imagery that are not forest. In other words, there are very few false negatives. Another promising aspect of the model is that where both models are successful at classifying forest, they are accurate at capturing the outline of plots of forest and capturing the entire plot.

There are many false negatives found in the classifications, where forest is not classified as anything. This is better than if there were many false positives by the model, however not ideal. The most likely cause of this is lack of adequate training. There is much variation between satellite imagery, which is likely what confused Peruforest 2 when it was rerun on the 2018 imagery (Appendix 2). The 2018 imagery appears to have the sun coming from the northeast, whereas in the 2022 imagery, the sun appears to be more directly overhead resulting in more shadows within stands of trees as opposed to off to the sides. The imagery also comes from two different satellites. There is also much variation between different plots of trees, as there is such a large elevation change within the valley which impacts the way that trees appear in the imagery. In order to combat all of these differences in appearance, as many years of imagery and iterations of the same area must be fully annotated and used as training samples for the model so that it is adequately prepared for all cases of pine and eucalyptus in imagery.

Conclusion

The most recently planted pine will not be ready for harvest for at least another 15 years. Due to this, there is ample time to continue improving on the performance and accuracy of this model to enable continuous monitoring of the area. This will be feasible by files included at the end of this document that instruct future students on how to continue training the model described in this paper.

Much more data will be needed to train the deep learning model on all types of situations that it can come across in imagery. The tree species being studied are simultaneously similar to one-another in their coloration and spectral signatures, while being diverse in their textures and shadowing within the species, depending on where it grows within the steep hillsides of the valley. Mass amounts of training can account for this.

A major conclusion from the interviews Keese conducted in August, 2022 was that there is a lack of communication between NGOs, locals, and government officials about reforestation projects. Understanding how forest transition is taking place at the community level can help these organizations in future monitoring and planning of forestry projects.

This study shows great promise as a future longitudinal study of land use and land cover change in Taucamarca. This research, coupled with Keese's working paper and ongoing research on the political ecology of forest transition in the Cusco region of Peru captures the critical importance of pine and eucalyptus to the dynamic of social and economic life in the Peruvian Andes.

Future Seniors: Link to Deep Learning Guide (Also Appendix 3)

Link to files (Imagery, Groundtruthing points, Deep Learning Packages, schema, etc.): Here

References

Alo, C. A., & Pontius, R. G. (2008). Identifying Systematic Land-Cover Transitions Using Remote Sensing and GIS: The Fate of Forests inside and outside Protected Areas of Southwestern Ghana. Environment and Planning B: Planning and Design, 35(2), 280–295. doi:10.1068/b32091

Bolyn, C., Lejeune, P., Michez, A. & Latte, N. (2022).Mapping tree species proportions from satellite imagery using spectral–spatial deep learning, Remote Sensing of Environment(280), 1-13. 10.1016/j.rse.2022.113205

Dickson, J. C. (1969). The Eucalypt in the Sierra of Southern Peru. Annals of the Association of American Geographers, 59(2), 294–307.

Guariguata, M., Arce, J., Ammour, T., & Capella, J. (2017). Las plantaciones forestales en Perú Reflexiones, estatus actual y perspectivas a futuro. Center for International Forestry Research (CIFOR). https://doi.org/10.17528/cifor/006461

Hamedianfar, A., Mohamedou, C., Kangas, A., Vauhkonen, J. (2022). Deep learning for forest inventory and planning: a critical review on the remote sensing approaches so far and prospects for further applications. Forestry: An International Journal of Forest Research, 95(4), 451–465. https://doi.org/10.1093/forestry/cpac002

Hartling, S., Sagan, V., Sidike, P., Maimaitijiang, & M., Carron, J. 2019. Urban Tree Species Classification Using a WorldView-2/3 and LiDAR Data Fusion Approach and Deep Learning. Sensors, 19(6). https://doi.org/10.3390/s19061284

Keese, J. (2022). Sabbatical Leave Application for Fall Quarter, 2022-23 AY.

Keese, J., Camacho, A., & Chavez, A. (2017). Follow-up study of improved cookstoves in the ‘

Cuzco region of Peru. Development in Practice, 27(1), 26–36. https://doi.org/10.1080/09614524.2017.1257565

Keese, J., Guzman Vazquez, C., O'Brien, M., & Richter, B. (2020). LPG Cookstove Use and Fuel Subsidies in the Cuzco Region of Peru. International Journal of Sustainable Development Research, 6(1), 14. https://doi.org/10.11648/j.ijsdr.20200601.13

Keese, J., Mastin, T., & Yun, D. (2007). Identifying and Assessing Tropical Montane Forests on the Eastern Flank of Abstract the Ecuadorian Andes. Journal of Latin American Geography, 6(1), 63–84.

Kussul, N., Lavreniuk, M., Skakun, S. and Shelestov, A. (2017). Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. IEEE Geoscience and Remote Sensing Letters, (14)5, 778-782. 10.1109/LGRS.2017.2681128.

Luzar, J. (2007). Luzar_Peru_2007 (1). Ethnobotany Research and Applications, 5, 85–93. https://ethnobotanyjournal.org/index.php/era/article/view/121

Mather, A. S., & Needle, C. L. (1998). The Forest Transition: A Theoretical Basis. Royal Geographical Society, 30(2), 117–124. https://about.jstor.org/terms

Mubin, N.A., Nadarajoo, E., Shafri, H.Z., & Hamedianfar, A. (2019) Young and mature oil palm tree detection and counting using convolutional neural network deep learning method. International Journal of Remote Sensing, (40)19, 7500-7515. 10.1080/01431161.2019.1569282

Nezami S., Khoramshahi E., Nevalainen O., Pölönen I., Honkavaara E. (2020) Tree Species Classification of Drone Hyperspectral and RGB Imagery with Deep Learning Convolutional Neural Networks. Remote Sensing, 12(7), 1-20. https://doi.org/10.3390/rs12071070

Raboin, M. L., & Posner, J. L. (2012). Pine or pasture? estimated costs and benefits of land use change in the peruvian andes. Mountain Research and Development, 32(2), 158–168. https://doi.org/10.1659/MRD-JOURNAL-D-10-00099.1

Ravikumar, A., Sears, R. R., Cronkleton, P., Menton, M., & Pérez-Ojeda del Arco, M. (2017). Is small-scale agriculture really the main driver of deforestation in the Peruvian Amazon? Moving beyond the prevailing narrative. In Conservation Letters (Vol. 10, Issue 2, pp. 170–177). Wiley-Blackwell. https://doi.org/10.1111/conl.12264

Rudel, T. K. (2009). Tree farms: Driving forces and regional patterns in the global expansion of forest plantations. Land Use Policy, 26(3), 545–550. https://doi.org/10.1016/j.landusepol.2008.08.003

Sefrin O., Riese F.M., Keller S. (2021) Deep Learning for Land Cover Change Detection. Remote Sensing. 13(78), 1-27. https://doi.org/10.3390/rs13010078

Tito, R., Salinas, N., Cosio, E. G., Boza Espinoza, T. E., Muñiz, J. G., Aragón, S., Nina, A., & Roman-Cuesta, R. M. (2022). Secondary forests in Peru: differential provision of ecosystem services compared to other post-deforestation forest transitions. Ecology and Society, 27(3). https://doi.org/10.5751/ES-13446-270312

Tovar, C., Seijmonsbergen, A. C., & Duivenvoorden, J. F. (2013a). Monitoring land use and land cover change in mountain regions: An example in the Jalca grasslands of the Peruvian Andes. Landscape and Urban Planning, 112(1), 40–49. https://doi.org/10.1016/j.landurbplan.2012.12.003

Weinstein, B.G., Marconi, S., Bohlman, S., Zare, A., White, E. (2019) Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. Remote Sensing, 11(11), 220-235. https://doi.org/10.3390/rs11111309

Zhang X., Han L., Han L., Zhu L. (2020) How Well Do Deep Learning-Based Methods for Land Cover Classification and Object Detection Perform on High Resolution Remote Sensing Imagery? Remote Sensing, 12(417), 1-29. https://doi.org/10.3390/rs12030417

Appendix

Peruforest1 Confusion Matrix

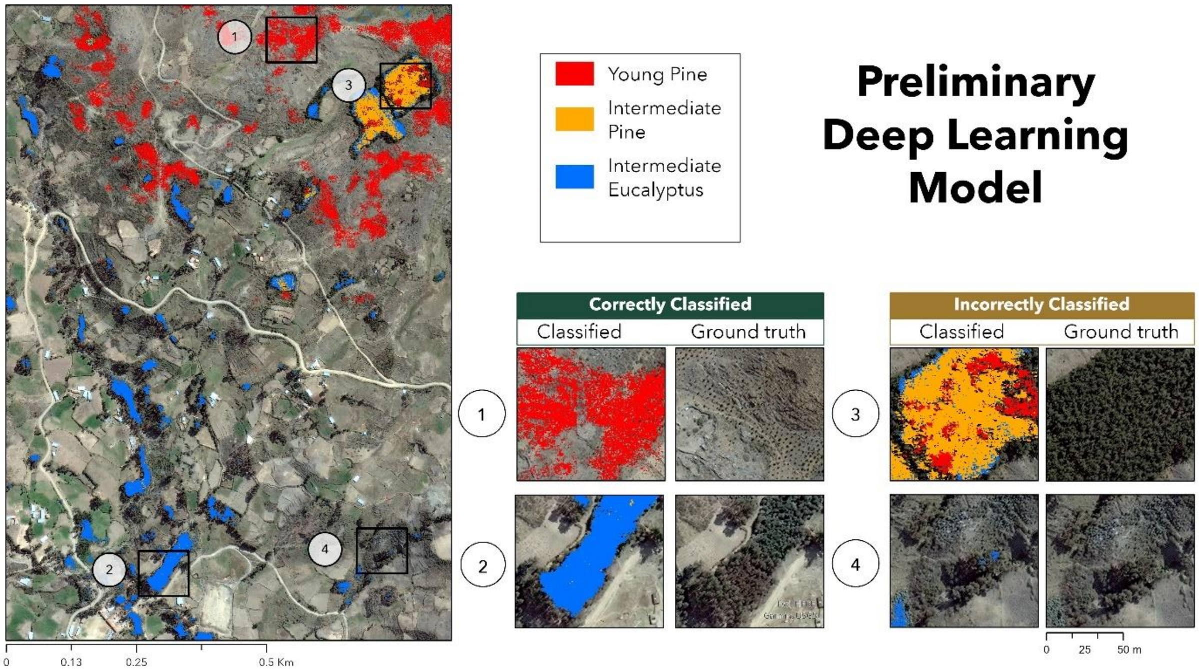
	Y. Pine	I. Pine	Y. Euc	I. Euc	Not Forest	Total	U_accuracy	Kappa
Y. Pine	6	0	1	0	1	8	0.75	0
I. Pine	0	1	0	2	0	3	0.3333333	0
Y. Euc	1	0	2	0	0	3	0.6666666	0
I. Euc	0	1	1	5	5	12	0.4166666	0
Not Forest	1	0	0	0	73	74	0.98648648	0
Total	8	2	4	7	79	100	0	0
P_accurac	0.75	0.5	0.5	0.7142857	0.0632911	0	0.9193599	0
Kappa	0	0	0	0	0		0	-0.488093

Kappa score of -48% and overall accuracy of 84%.

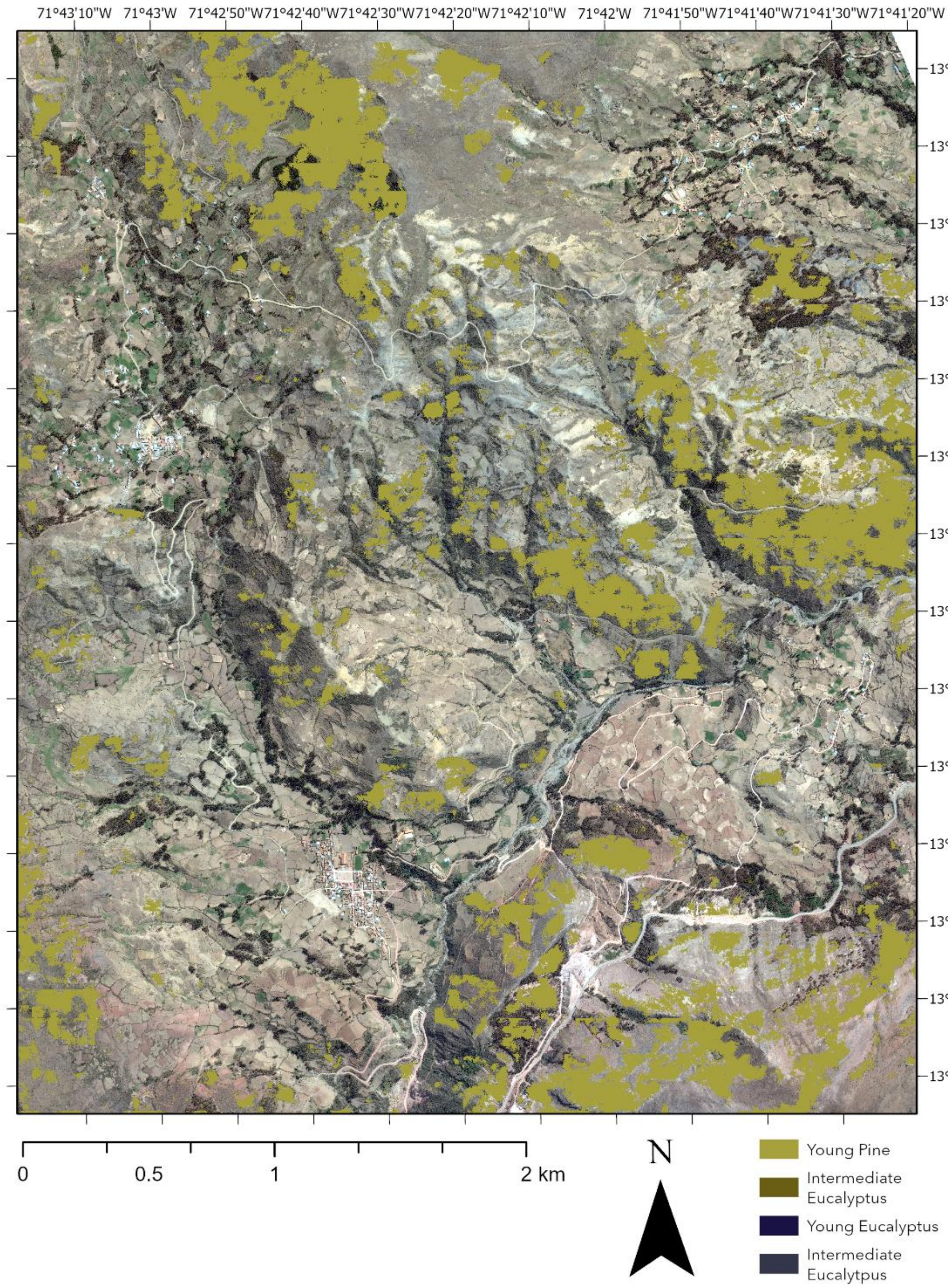
Peruforest 2 Confusion Matrix

	Y. Pine	I. Pine	Y. Euc	I. Euc	Not Forest	Total	U_accura	Kappa
Y. Pine	1	1	0	0	5	7	0.142857	0
I. Pine	0	0	0	0	1	1	0	0
Y. Euc	0	0	0	0	7	7	0	0
I. Euc	0	0	0	0	2	2	0	0
NotForest	1	1	0	0	81	83	0.975903	0
Total	2	2	0	0	96	100	0	0
P_accura	0.5	0	0	0	0.0208333	0	0.843235	0
Kappa	0	0	0	0	0	0	0	-0.148215

Kappa score of –14% and overall accuracy of 84%



Appendix 1 Preliminary Peruforest model results with partial training.

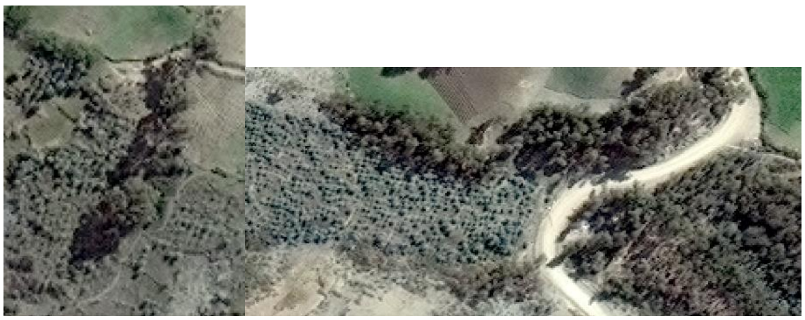


Appendix 2 Peruforest2 run on the 2018 imagery. Only young pine in classified in the valley, and very inaccurately compared to the other times the models are run.

Appendix 3 Deep Learning Guide for Future Students

This document was created by Sadie Calhoun in order to pass on the monitoring of forest transition in Taucamarca by training a deep learning model to differentiate between pine and eucalyptus. Questions can be directed to Sadie at sadcalhoun@gmail.com.

Eucalyptus:



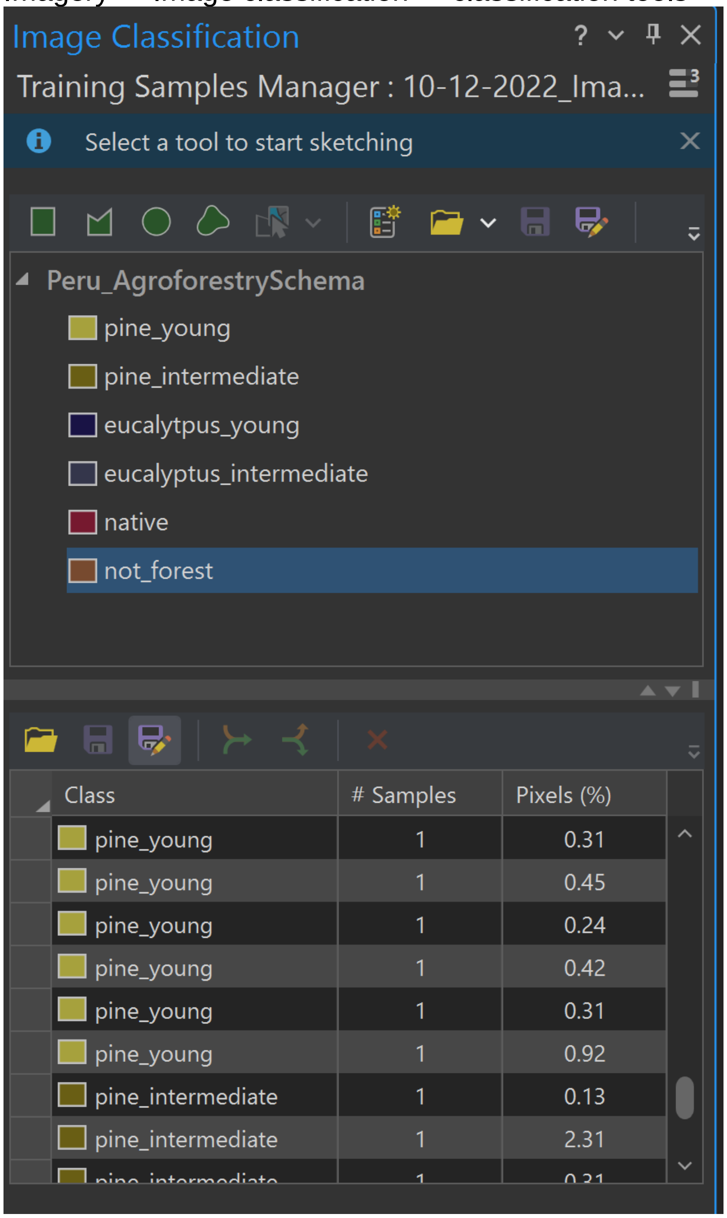
Young eucalyptus appears more turquoise in imagery. It also is rounder and looks ball like. Intermediate eucalyptus takes on more of a grey-green color.

Pine:

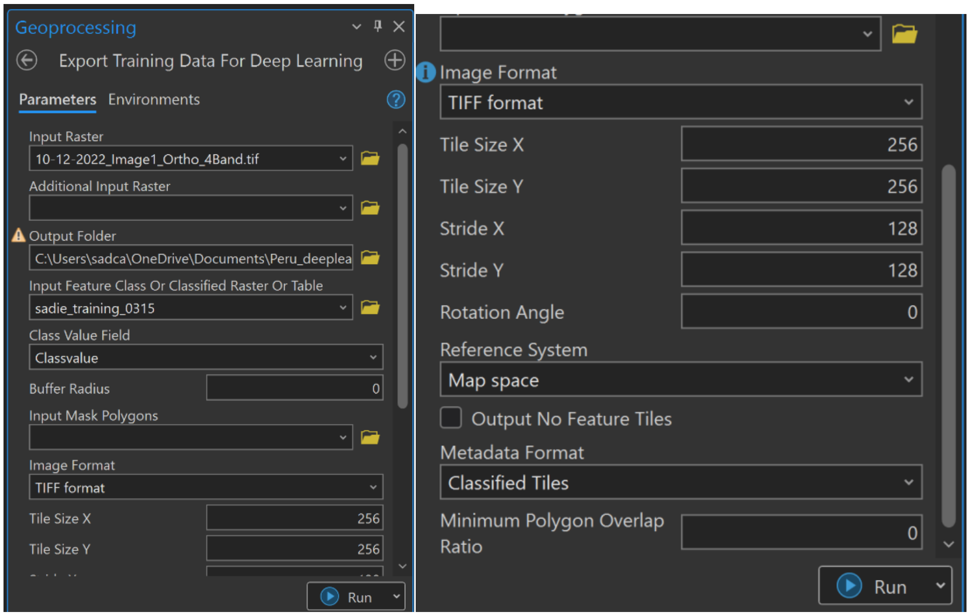
Young pine is planted in clear rows. It appears as small to medium size dots on the imagery. Intermediate pine can also sometimes be seens in the clear rows, but as they fill in, it becomes an evergreen color.

How to Continue Training Deep Learning Models!

1. Imagery -> Image classification -> classification tools -> Training Samples Manager



2. Open training scheme (link to onedrive w schema above)
3. Draw polygons around areas of forest that you are positive about (refer to images above)
4. Save your training samples as a feature class
5. Add the feature class to your map
6. Open tool Export Training Data for Deep Learning



*Input Raster = Satellite imagery you want the samples to consist of

*Output Folder = Best to save to local geodatabase folder. A good naming convention (the one that we used) is Name_samples_date

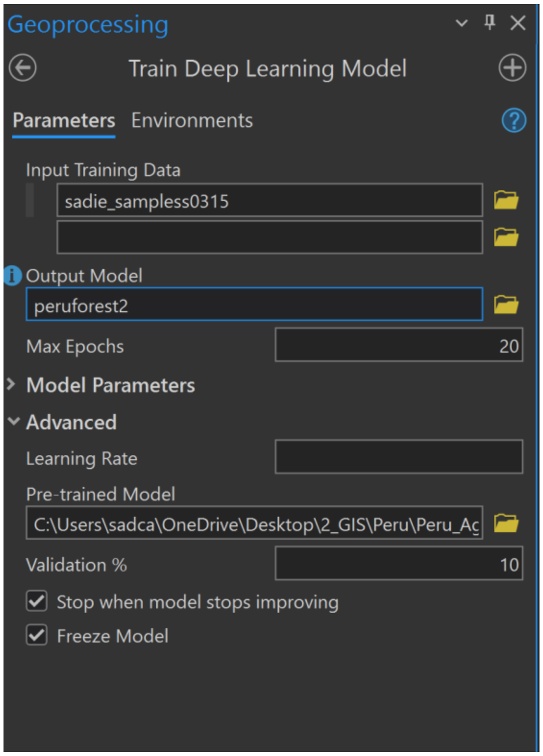
*Class Value Field = Classvalue

***Image Format = TIFF format

***Metadata Format = Classified Tiles

***These are key to change, the default will be PASCAL Visual Object, and you won't be able to train the .dlpk if you do not change this.

7. Open the Geoprocessing tool [Train Deep Learning Model](). This requires high computing power, so it best done on the Lab Computer if you don't have a strong GPU

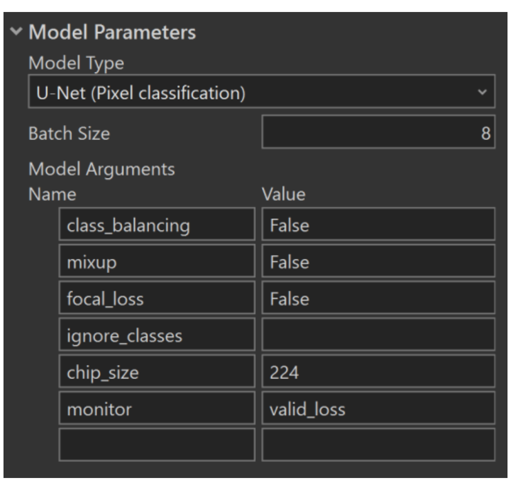


*Input Training Data = your files exported in step 6.

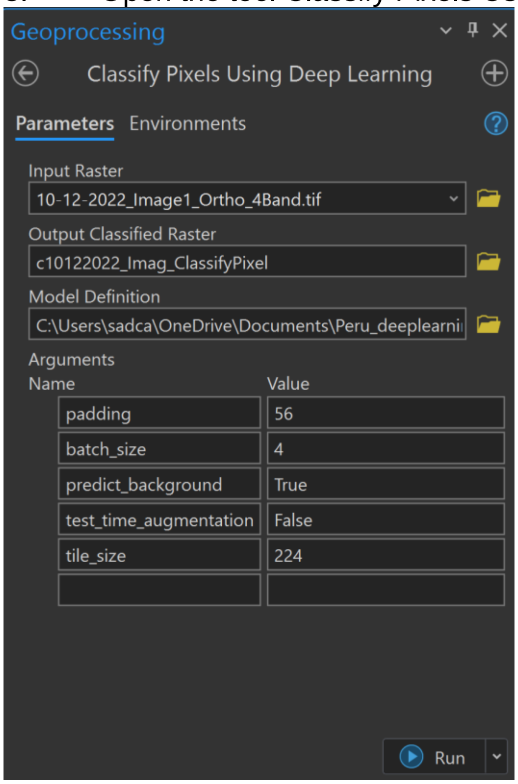
*Output Model = peruforestx, where x is whatever version of the model you are on. This output model is best saved in the geodatabase for the project, with a local copy saved as well.

*Pre-trained Model = the most recently trained .dlpk (should theoretically always be peruforest(x-1)). Once you put this in, the Model Parameters will disappear because it will use the parameters from the pre-trained model. These parameters are below:

***Make sure that your imagery has the required number of pyramids (6)



8. Open the tool Classify Pixels Using Deep Learning



*Input Raster = the imagery you want to run the .dlpk on

*Model Definition = the .dlpk file generated from step 7.

Imagery trained on so far

	Sept. 6 2018	July 9 and Oct. 12 2022
Satellite	Worldview-3	Jilin-1
Resolution	50 cm	50 cm
Bands	4-band	4-band