THESIS

TOWARDS RECONSTRUCTING CITIES WITH AI:

A NOVEL MACHINE LEARNING APPROACH FOR AUTOMATED ARCHAEOLOGICAL SURVEYING AND PRESERVATION BY LEARNING CANOPY STRUCTURES.

Submitted by

Nicholas Redford

Department of Anthropology and Geography

In partial fulfillment of the requirements

For the Degree of Master of Arts

Colorado State University

Fort Collins, Colorado

Summer 2025

Master's Committee:

Advisor: Chris Fisher

Elizabeth Tulanowski
Stephen J. Leisz

ABSTRACT

TOWARDS RECONSTRUCTING CITIES WITH AI: A NOVEL MACHINE LEARNING APPROACH FOR AUTOMATED ARCHAEOLOGICAL SURVEYING AND PRESERVATION BY LEARNING CANOPY STRUCTURES.

Archaeological landscapes face increasing threats from climate change and human activity, necessitating scalable methods for site detection. LiDAR has revolutionized archaeological surveys by revealing features beneath dense vegetation, but manual interpretation remains labor-intensive. This study introduces an Automated Archaeological Survey Method (AASM) that utilizes machine learning to analyze tree canopy structures as proxies for underlying archaeological features. Using a LiDAR-derived digital canopy model (DCM), the Bi-path Ensemble (BPE) model predicts the locations of the land-use typologies of “public” and “private” space at Angamuco. The model evaluation shows moderate to high agreement between predicted and true labels, particularly for dense public and un-terraced private spaces. These results suggest that vegetation patterns can serve as reliable indicators of past human activity, offering a scalable approach for prioritizing areas for archaeological surveys. By integrating creative computational methods with remote sensing data, this study advances the use of machine learning in archaeological landscape reconstruction.

TABLE OF CONTENTS

Chapter 1: Introduction

Archaeological sites worldwide face increasing threats from both climate change and human development. In 2019, the International Council on Monuments and Sites (ICOMOS) declared a state of emergency, warning that global climate changes were accelerating the destruction of cultural heritage (ICOMOS, 2019). McGovern (2018) described this crisis as akin to "burning libraries," where vast, irreplaceable records of human history—documented through material culture—are being erased. Rising sea levels, deforestation, agricultural expansion, and infrastructure projects continue to endanger archaeological landscapes, often before they are even documented (McGovern, 2018). Hesse (2009) emphasized that many of these undetected archaeological features are in forested regions, obscured by dense vegetation that complicates archaeological survey methods (Hesse, 2009). Archaeologists must develop scalable methods for surveying large arboreal regions to identify and study unrecorded sites before they are permanently damaged by climate change or urban development.

Traditional ground-based archaeological survey methods struggle to overcome challenges in regions with dense vegetation, rugged terrain, and limited accessibility. The emergence of LiDAR-based survey techniques in the 21st century revolutionized archaeological research in forested regions by revealing hidden landscapes beneath forest canopies (Chase et al., 2012). However, this technology introduced new challenges—LiDAR scans generate large datasets that archaeologists must manually analyze to identify cultural features. Consequently, there is a growing need for automated methods to process these datasets efficiently, scaling archaeological expertise over entire landscapes and narrowing areas to prioritize areas for detailed investigation.

Automated Archaeological Survey Methods (AASMs) have emerged as an important response to this challenge, leveraging remote sensing and computational tools to detect potential archaeological features across vast and complex environments. The research and the purpose of this thesis arrives in this paradigm. I hope to build upon previous research that significantly links tree canopy structures to the locations of archaeological material (Hightower et al., 2014; Vaughn & Crawford, 2009). In this thesis, I present a novel AASM method that leverages tree canopy structures and machine learning to predict the locations of archaeological phenomena (mainly structural features such as terraces, house mounds, roads, ceremonial structures etc.). Beyond identifying likely archaeological sites, this model also predicts broader land-use typologies, offering insights into regional settlement patterns. By providing a landscape-scale perspective that outlines more general urban patterns and contours the distribution of land use, I hope this approach can complement and contextualize more precise AASMs that extract individual features.

Research Objectives and Hypothesis

This thesis investigates whether tree canopy structure can be a reliable proxy for reconstructing archaeological landscapes. Given the demonstrated relationship between vegetation patterns and archaeological phenomena, I seek to determine if features engineered from a LiDAR-derived digital canopy model (DCM) can predict the locations of specific assemblages of archaeological features.

The key goals of this thesis are as follows:

1. Develop an AASM that analyzes canopy structures to identify potential archaeological features.

2. Overcome challenges in archaeological data by developing a predictive model capable of handling data with subtle, non-linear, and multiplex relationships.

3. Evaluate model performance using statistical metrics (Balanced Accuracy, AUC, MCC, Kappa, and percentage change to control groups) and visual validation to assess its ability to distinguish between areas with and without the relevant archaeological features.

The central hypothesis of this research is that patterns in tree canopy structure can effectively identify specific typologies of areas associated with past human activity. If successful, this approach could be used to automatically reconstruct archaeological landscapes, and guide decisions to prioritize certain areas for high-resolution archaeological feature extraction (AFE) and ground survey efforts.

Roadmap

Chapter 2: Background – This chapter provides the necessary foundation for understanding the study's methodological framework. It begins with a historical overview of remote sensing in archaeology, tracing its evolution from early aerial photography to modern LiDAR and machine learning applications. It then explores the development of Automated Archaeological Survey Methods (AASMs), situating this study within broader trends in predictive modeling and feature extraction. Finally, it presents relevant background on Angamuco, including its archaeological history, known site distributions, and ecological characteristics, which are crucial for understanding how canopy patterns might reflect past human modifications.

Chapter 3: Methods – This chapter details the study's methodological approach. It first outlines the process of label construction and feature selection, focusing on Dense Public Space, Unterraced Private Space, and Rural Terrace as the primary labels of interest. An ANOVA

analysis is used to confirm if significant distinction between created labels exists. Then PerMANOVA analysis is conducted to evaluate which features show the strongest differentiation between archaeological and non-archaeological areas. The second half of the chapter describes the computational framework of the most successful model architecture, nicknamed the BPE, including the machine learning algorithms used, feature engineering strategies, and the final architecture of the predictive model.

Chapter 4: Results and Discussion – This chapter presents the results of the study, evaluating the model's performance using metrics such as AUC, MCC, F1, and balanced accuracy. It discusses how well the BPE model predicts each label, considering both statistical and machine learning performance. The discussion further examines the implications of the model's successes and limitations, particularly in relation to previous work on AASMs and the broader goal of developing scalable methods for archaeological survey.

Chapter 5: Conclusion – The final chapter synthesizes the study's key findings, reflecting on the strengths and limitations of the approach. It discusses how the insights gained from canopy structure analysis can inform future archaeological work and highlights potential directions for refining predictive models through the integration of additional contextual variables.

Central Argument

This study demonstrates that a complex machine learning model can effectively predict the locations of archaeological features in arboreal settings by analyzing patterns in tree canopy structures. Through iterative refinement and model experimentation, I developed a unique multimodal architecture, nicknamed the Bi-path Ensemble (BPE), designed to accommodate the complex and variable relationships between the canopy structure and archaeological phenomena.

The BPE pipes data through two pathways and averages prediction scores from one hundred eighty iterations. The BPE consistently outperformed control group predictions. Furthermore, evaluation statistics on test sets showed moderate to high agreement between predicted and true labels when identifying dense public (urban) space and un-terraced private (urban) space. Visual validation of a DEM suggests that the BPE achieves decent accuracy in predicting exact labels both within and beyond the training region; and exhibits relatively high accuracy when considering exact or related labels within and beyond the training region. These results support the hypothesis that certain urban and settlement patterns leave predictable signatures in vegetation. Ultimately, this study highlights the potential of leveraging tree canopies to develop scalable methods to automatically reconstruct archaeological landscapes in arboreal settings, narrowing down points of interest for more granular study.

Chapter 2: Background

This chapter establishes the historical, environmental, and methodological context that informs the predictive modeling approach used in this study.

First, I explore the relevant historical and topographic context of Angamuco. I draw attention to the city's decentralized urban layout, where archaeological features are dispersed rather than concentrated in a central core. This spatial organization, shaped by a complex road network and prolonged occupation history, necessitates a predictive model that can accommodate diverse and interwoven architectural typologies. The second section examines the ecological setting, emphasizing how Angamuco's rugged volcanic terrain and dense oak canopy influence feature visibility. The next section reviews remote sensing applications, particularly LiDAR-based methods that have proven effective in detecting archaeological features beneath dense vegetation. Finally, I discuss machine learning applications in archaeology, highlighting both its potential and limitations. By integrating historical, ecological, and methodological insights, this chapter lays the foundation for a predictive modeling framework that moves beyond conventional site-based approaches. The findings presented here underscore the need for a model that can accommodate Angamuco's spatial complexity and leverage environmental signals to enhance feature detection.

History and Topography

Until 2009, most of the evidence for the cultural history of the Pátzcuaro Basin came from ethnohistoric sources, namely the Relación de Michoacán. The Relación, along with

excavations at Tzintzuntzan, painted the basin as a fairly rural region until the establishment of the Purépecha Empire in the Postclassic period (Pollard, 1980, 2008; Pollard et al., 2003). In 2010, LiDAR scans of a dried Pleistocene lava plateau (Malpaís de Zacapu) in the basin revealed a massive city buried under a thick forest canopy. The city, named Angamuco, reimagined the history of Michoacán by demonstrating the emergence of dense urbanism in the Pátzcuaro Basin long before the establishment of a Purépecha state. Angamuco was a large city by Postclassic (~1000-1500 CE) standards. Furthermore, the city was very dense, containing an estimated 20,000+ structures. Angamuco maintained a significantly high population from at least 1000 CE until the early modern period (Fisher & Leisz, 2013). In the 14th century, Angamuco became incorporated into the emerging Purépecha Empire. This political shift brought about architectural changes to the site, namely the Yácata (keyhole) style pyramid that became a staple of Purépecha influence (Cohen, 2016).

Figure 1: Hillshade of The Malpais of Zacapu (Angamuco). Created with ArcGIS pro.

An object-based image analysis (OBIA) conducted by Urquhart (2015) of the urban organization of Angamuco demonstrated that buildings grouped together into distinct complejos (neighborhoods) and larger districts (Urquhart, 2015). This urban planning logic reflects a typical

nested social organization system for Postclassic Central Mexico. Ethnohistoric evidence has long demonstrated that ancestral Nahua communities stratified into nested social and labor units called altepetl and tlaxilacalli, which affected both social organization and urban design (Johnson, 2018; Lockhart, 1992). More recent ethnohistorical investigations, combined with a widescale analysis of Angamuco's road networks and urban organization, suggest that similar nested social units existed at Angamuco, termed the Irieta system (Solinis-Casparius, 2019; Urquhart, 2015). These units are visible in the built environment as distinct areas of dense architecture, bounded by a road network.

Solinis-Casparius (2019) conducted perhaps the most comprehensive study of Angamuco's urban design by focusing on its extensive road system. His analysis highlights that roads were integral to structuring movement, infrastructure, and social interaction within the city. Using a combination of LiDAR, GIS spatial analysis, and excavation, Solinis-Casparius analyzed roads based on morphology and studied how they were created, maintained, and modified over time. His findings suggest that the road system defined social units at multiple scales, with complejos forming the smallest units, grouped into neighborhoods, and further aggregated into districts. Solinis-Casparius also identified key distinctions between public and private space at Angamuco. He found that roads played a crucial role in defining access to different areas of the city. Primary roads connected major entrances and high-traffic nodes, facilitating public interaction, while smaller roads and pathways led to more enclosed areas associated with private residential spaces. This distinction aligns with broader anthropological theories that urban layouts configure social interactions by determining accessibility and movement. In Angamuco, the redundancy of road networks and the absence of a single central space suggest that social organization was highly localized, with complejos functioning as semi-autonomous units within

a larger urban fabric (Solinis-Casparius, 2019). The long urban history of Angamuco, lack of central urban design, clustered Irieta urbanism, and the presence of external architectural styles indicate a multiplex of architectural and archaeological feature typologies.

Ecology and Environmental Context

The Malpaís de Zacapu is a Late Holocene volcanic plateau that formed through four separate eruptions beginning around 1450 BCE, with repose periods ranging between 400 and 1000 years. This volcanic activity created an elevated, rugged terrain that provided both natural construction materials and strategic advantages, such as visibility over the basin to monitor approaching groups. The site's position between the forests in the hinterland and the lakeshore allowed its inhabitants to exploit diverse ecological resources (Reyes-Guzmán et al., 2023).

Dorison (2022) conducted an ethnobotanical analysis of the Malpaís de Zacapu, arguing that ancestral Purépecha held advanced knowledge of soils, which allowed them to take advantage of the benefits of a volcanic ecological zone. The loose volcanic soils were easy to work and retained water efficiently due to their porosity, while native oak trees played a crucial role in nutrient recycling. These environmental conditions would have supported agriculture and habitation despite the seemingly inhospitable terrain. Archaeological surveys further highlight this adaptation, showing that terraces are the most common archaeological feature at Angamuco, comprising over 70% of nearly 5,000 recorded agrarian structures which appear in both urban and rural contexts. Their construction varied based on local slope, soil, and water dynamics. Additionally, large contiguous parcel systems on flatter lava flows suggest widespread land management through systematic stone clearance and accumulation. This evidence indicates that the terraces at Angamuco evolved through multiple construction phases and were integral to its landscape (Dorison, 2022).

The malpaís of Zacapu is covered in a high-elevation oak (Quercus) forest. The Inventario Nacional Forestal y de Suelos (INFyS) of Mexico categorizes most of the region around the Pátzcuaro Basin as a *bosque latifoliado*, or broadleaf tree biome, based on ground data collected between 2015 and 2020 (INFyS, 2015). Valderrama et al. (2014) similarly defined the area around Angamuco as a deciduous forest based on satellite data from 2000 (Valderrama et al., 2014). For latifoliado forests at 2000+ meters above sea level, the INFyS estimates that the canopy is composed predominantly of Quercus species, with an average tree height ranging from 4 to 7 meters (INFyS, 2015). The reports from INFyS and Valderrama et al is further supported by Dorison's analysis which notes that oak forests typically dominate the malpais in the Pátzcuraro basin (Dorison, 2022).

Historical and Ecological Implications for Predictive Modeling

The ecological and historical context of Angamuco presents unique challenges for predictive modeling. The site's archaeological features are diverse, interspersed, and have been constructed and modified over a long period of habitation, which may make precise typological predictions difficult. Unlike sites with centralized cores, Angamuco’s decentralized layout consists of dispersed public, private, and agrarian spaces with varying architectural forms. This complexity requires a model capable of detecting nuanced spatial relationships rather than relying on singular patterns. Additionally, the homogeneity of the oak-dominated canopy means that predictive modeling cannot rely on tree species as a proxy for archaeological features. Instead, structural variations in the canopy—such as differences in height, density, and spatial arrangement—may provide more reliable indicators.

Fortunately, extensive spatial analyses that utilize the deep archives of remotely sensed data at Angamuco have demonstrated that archaeological features can be successfully extracted using GIS and machine learning methodologies. These findings provide an optimistic outlook, suggesting that despite the site's complexities, predictive modeling can overcome these challenges and yield meaningful insights.

Remote Sensing and Spatial Analysis

Remote sensing has fundamentally changed archaeological survey methods by enabling large-scale site detection and spatial analysis of entire regions. While aerial photography provided the first remote sensing applications in archaeology—such as British military surveyors photographing Mesopotamian sites in 1919—it remained costly and limited to open landscapes (Leisz, 2013). The launch of the Landsat satellite in 1972 made multispectral remote sensing data widely available, but archaeologists initially struggled to use it due to high processing costs and technological limitations (Ebert, 1984; Wiseman & El-Baz, 2007).

The early 2000s saw the growing adoption of Geographic Information Systems (GIS) which enabled archaeologists to analyze spatial data quantitatively and generate useful visualizations. Estrada-Belli and Koch (2007) demonstrated GIS applications by modeling Maya road networks and correlating site distributions with environmental factors (Estrada-Belli & Koch, 2007). These studies showcased the potential of remote sensing for spatial analysis but were still constrained by the resolution of available satellite imagery.

The increased availability of Light Detection and Ranging (LiDAR) later in the decade offered new opportunities for spatial analysis. Chase et al. (2012) argued that LiDAR technology

could revolutionize archaeological praxis—potentially as much as radiocarbon dating. They specifically cited the ability of LiDAR to pierce forest canopy and generate unobfuscated visualizations, permitting archaeologists to survey previously unreachable arboreal sites. Forested archaeological sites create numerous practical challenges for surveying and spatially contextualizing archaeological sites. Furthermore, these sites cannot be comprehensively studied with satellite imagery, considering tree canopies obscure structures underneath. With LiDAR-generated digital elevation models (DEMs), archaeologists can finally spatially contextualize and digitize the once-problematic forested sites (Chase et al., 2012).

DEMs have proven to be highly effective representations of cultural landscapes. Reese-Taylor et al. (2016) demonstrated that while some vegetation species can introduce errors—sometimes causing cultural features to be obscured or misrepresented—DEMs derived from LiDAR data are generally reliable for identifying archaeological features. Their research emphasized that, despite occasional inaccuracies caused by vegetation, the overall accuracy of DEMs in representing cultural landscapes is high, making them valuable tools for archaeological surveys and feature detection (Reese-Taylor et al., 2016).

Beyond methodological accuracy, LiDAR has also altered archaeological interpretations on a broader scale. Mlekuž (2013) emphasized that LiDAR does more than just detect sites—it reshapes our perception of them. Because modifications to landscapes often extend beyond the boundaries of individual sites, LiDAR reveals the interconnected nature of past human activities across entire regions (Mlekuž, 2013). This perspective challenges traditional site-based archaeological interpretations, reframing cultural heritage as an integrated landscape rather than a collection of isolated features. However, leveraging LiDAR to analyze archaeology at a landscape scale presents a practical challenge due to the sheer volume of data. To address this,

automated feature extraction (AFE) techniques have emerged to identify points of interest across large datasets. These methods have become crucial for translating LiDAR's regional perspective into actionable archaeological insights

AFE approaches have been widely employed to detect archaeological features in various geographic and environmental contexts. Hesse (2009) was among the first to demonstrate the potential of automated methods for detecting topographic anomalies, such as burial mounds, agricultural terraces, and sunken roads, in Germany. His work highlighted the importance of computational models in overcoming limitations associated with traditional visual interpretation of remote sensing data (Hesse, 2009). Similarly, Freeland et al. (2016) applied AFE techniques in the Kingdom of Tonga to identify mounds, comparing results from object-based image analysis (OBIA) and a modified pit-detection algorithm. Their study demonstrated the sensitivity of automated detection methods while also revealing high rates of false positives, emphasizing the need for complementary verification strategies (Freeland et al., 2016).

Successful AFEs and AASMs have been created at Angamuco. Harris (2019) developed a methodology using GIS-based topographic manipulation to extract the built environment from the site's rugged malpaís terrain. His approach focused on identifying potential archaeological features by separating cultural modifications from natural topography through value identification and analysis preparation. The final dataset consisted of 87,407 features, which were then analyzed using density measurements and classification techniques based on the Thinness Ratio to conduct OBIA at an individual feature level. By applying automated methods to a complex landscape, Harris's work exemplifies how AFE can efficiently process vast LiDAR datasets, transforming raw spatial information into a more meaningful archaeological visualization (Harris, 2019).

Spatial analysis of remotely sensed data has allowed archaeologists to study not only archaeological materials themselves but also how past human activities shaped and interacted with the surrounding environment. Rather than focusing solely on detecting structures, researchers have increasingly examined how archaeological features influence ecological patterns, such as vegetation growth and canopy structure. Initially, archaeological applications of LiDAR focused on stripping away vegetation to expose cultural features. However, rather than viewing tree canopies as obstacles, archaeologists can also leverage them as sources of information. Several studies have demonstrated that vegetation structure and composition correlate with underlying archaeological materials. Hightower et al. (2014) used LiDAR-derived digital canopy models (DCMs) to show significant variations in tree canopy dimensions over archaeological terraces in Caracol, Belize (Hightower et al., 2014). Vaughn and Crawford (2009) integrated NDVI and greenness indices from spectral data with topographic and hydrological features to predict Maya settlement locations in Northwestern Belize, finding that increased vegetation density corresponded with abandoned settlements due to moisture-retaining depressions from past limestone extraction (Vaughn & Crawford, 2009).

These studies highlight that remotely sensed vegetation data—whether from LiDAR or spectral imaging—can reveal meaningful archaeological patterns. By incorporating canopy structure alongside topographic information, remote sensing approaches can enhance archaeological predictive models, extending beyond simple terrain analysis.

The application of remote sensing in archaeology has expanded beyond identifying structures to understanding how past human activities influenced the environment. Studies demonstrating correlations between canopy structure and archaeological materials, such as those by Hightower et al. (2014) and Vaughn and Crawford (2009), played a major role in shaping this

research. Their findings suggest that vegetation patterns can serve as key indicators of underlying archaeological features. This thesis builds on that foundation, integrating both LiDAR and spectral data to explore how canopy structure can be leveraged in a machine learning integrated AASM.

Machine Learning

Machine learning (ML) is a computational approach that enables systems to learn from data and adapt without explicit programming. Emerging in the mid-20th century with rudimentary game-playing algorithms, ML has evolved into a cornerstone technology used in fields ranging from medical diagnostics to security. Unlike traditional statistical models, ML is "soft-coded," meaning models refine themselves by recognizing patterns in data rather than relying on predefined rules. This adaptability connects ML not only to mathematics but also to cognitive science, information theory, and philosophy (El Naqa & Murphy, 2015). The rapid development of ML is intertwined with advancements in data storage and computational power. As datasets expanded, traditional analytical methods struggled with complexity, necessitating ML models to process and simplify large-scale information (Jordan & Mitchell, 2015). While ML was initially developed to handle "big data," its applications have extended to diverse fields, including archaeology.

ML in Archaeology: AASM that Classify land use

Archaeology presents an unconventional challenge for ML due to its reliance on "slow data"—fragmented datasets with inherent biases. However, the rise of remote sensing has provided archaeologists with extensive geospatial datasets, making ML a viable tool for site detection and classification (Bickler, 2021). Classification models, which assign discrete labels to data based on learned relationships, are particularly relevant for automating archaeological

surveys. Archaeological classification is not solely a machine learning task—it is a common analytical process that can be enhanced by computational methods such as GIS. While classification for spatial datasets in archaeology has traditionally relied on expert interpretation, machine learning can augment this process by identifying patterns across vast datasets, refining automated archaeological survey methods (AASMs), and improving feature extraction from remote sensing data.

While critics argue that ML lacks the contextual understanding necessary to reliably classify archaeological features, recent breakthroughs suggest otherwise (Davis, 2021). These current studies into ML-based AASMs can be divided into two methods of classification:

Per-Pixel Classification:

Early ML-based AASMs used satellite-derived spectral data to classify pixels based on their spectral signatures. These methods excel in broad-scale land-use categorization but struggle with finer archaeological features due to resolution limitations (Myint et al., 2011). Per-pixel classifications can be further categorized into two types. "Direct" classification methods identify archaeological materials based on their inherent characteristics, such as spectral signatures. For instance, Ben-Romdhane et al. (2023) employed a direct per-pixel approach, classifying pixels based on the spectral signature of the archaeological materials themselves, using a neural network to analyze hyperspectral data from open-air sites in the UAE (Ben-Romdhane et al., 2023). In contrast, "indirect" pixel-based methods rely on environmental variables as proxies for archaeological phenomena. Vinicius-Periato et al. (2023) utilized an indirect Per-Pixel AASM approach by analyzing hyperspectral data to detect tree species exclusive to anthropogenic soils in the Amazon, indirectly identifying locations of archaeological sites. Notably, this contrasts

with Ben-Romdhane's method because it classified the spectral signatures of pixels representing an environmental proxy rather than the archaeological material itself (Peripato et al., 2023).

Maximum Entropy (MaxEnt) algorithms represent one of the most popular indirect per-pixel method. MaxEnt correlates environmental variables with known site locations to predict settlement patterns. This approach has proven valuable in lower-resolution contexts, as demonstrated by Howey et al. (2016) in their study of medieval subterranean food storage in Michigan (Howey et al., 2016). Rafuse et al. (2021) applied a similar method to hunter-gatherer sites in Argentina, using model insights to infer settlement preferences (Rafuse, 2021).

Object-Based Image Analysis (OBIA):

OBIA classification methods classify pixels within spatially defined objects, often improving contextual recognition (Dey et al., 2011). OBIA methods such as those used by Hesse (2009) and Harris (2019) can be enhanced by machine learning components. Machine learning algorithms in these contexts not only extract points of interest but can classify them as specific, defined archeological typologies. Convolutional neural networks (CNNs) have shown promise in their ability to automatically extract archaeological features from LiDAR-derived DEMs. CNNs are a type of neural network that excel in feature detection by learning hierarchical spatial patterns. Vaart & Lambers (2019) used a CNN to successfully identify and classify large-scale features such as Celtic fields but struggled with smaller elements like kilns (Vaart & Lambers, 2019). Similarly, Guyot et al. (2021) used CNNs to uncover previously undocumented mounds and pits in Southern Brittany, France. An OBIA ML classification method has been shown to be effective at Angamuco. Building upon Harris's (2019) AFE methodology, Albrecht et al. (2019) introduced a deep-learning component to feature extraction at Angamuco. Using a specialized variety of CNN called a visual geometry group (VGG) artificial network, their approach sought

to automate classification at a more granular level, potentially distinguishing between different building types. Their model achieved a relatively high agreement (AUC of 0.82) when identifying house structures, demonstrating that deep learning could enhance the precision of AFE for complex archaeological sites. The introduction of machine learning marked a significant step forward in scaling archaeological expertise, reducing the need for manual interpretation, and refining previous GIS-based methods (Albrecht et al., 2019).

Enduring Challenges

Despite promising results, AASMs consistently face high false-positive rates. Critics argue that archaeological materials are too inconsistent and environmentally embedded for ML to provide reliable predictions. However, ongoing refinements—such as hybrid models combining multiple data sources—continue to improve accuracy (Davis et al., 2019; Guyot et al., 2021).

Chapter 2 Conclusion

The history, urban organization, environmental context, and prior applications of remote sensing and machine learning at Angamuco all inform the approach taken in this study. The city's decentralized layout, as identified by Solinis-Casparius (2019), suggests that archaeological features are not concentrated in a central core but rather dispersed within a complex road network. This unique spatial organization necessitates a modeling approach that can detect patterns beyond traditional urban planning structures. The long occupation history and architectural diversity further complicate predictive efforts, as different feature types may exhibit varying levels of preservation and visibility.

Ecologically, the site's rugged volcanic terrain and dense oak canopy present challenges for feature detection. Unlike other regions where vegetation signals serve as reliable proxies for

human activity, the relative homogeneity of Angamuco's forest cover limits the effectiveness of methods that rely on spectral differentiation. However, topographic variation, soil composition, and hydrological patterns may still offer valuable environmental cues for identifying anthropogenic modifications to the landscape.

Advances in remote sensing and machine learning provide promising solutions to these challenges. Previous studies have demonstrated that LiDAR-derived data, when combined with predictive modeling, can effectively detect archaeological features despite dense vegetation cover. Methods such as automated feature extraction (Hesse 2009, Freeland et al., 2016; Harris 2019) have shown that terrain-based anomaly detection can uncover hidden cultural landscapes. Additionally, machine learning approaches, such as those employed by Albrecht et al. (2019) at Angamuco, have highlighted the potential for predictive classification, though with varying levels of accuracy. These insights inform the methodological framework of this study, which aims to supplement current research into predictive modeling for archaeology by incorporating a holistic set of environmental signals to detect archaeological phenomena.

Chapter 3: Methods

This chapter outlines the methodological framework used to develop a predictive model for identifying archaeological typologies at Angamuco. The process begins with an overview of raw data collection, detailing the remote sensing and environmental datasets that serve as inputs for analysis. From there, the chapter describes the creation of labels, which define the archaeological categories of interest and serve as training data for supervised machine learning models. These labels were evaluated through statistical analyses, including ANOVA and PerMANOVA, to determine whether meaningful distinctions exist among them and to assess their explanatory power over a suite of environmental variables. Based on these results, three key label categories—Dense Public Space, Unterraced Private Space, and Rural Terrace—were selected for predictive modeling.

Following label selection, the chapter introduces various machine learning algorithms considered for this task, ultimately identifying gradient boosting and neural networks as the most promising approaches. However, given challenges related to sensitivity and overfitting, a multimodal strategy was developed to balance these limitations. The chapter then details the statistical evaluation techniques used to refine model architecture iteratively, ensuring robustness in predictive performance. Finally, the methodology concludes with an overview of the final multimodal approach, the BPE, which integrates canopy structure and topographic data to infer the spatial distribution of certain archaeological phenomena.

Data Collection

Studies like Reese-Taylor et al. (2016) and Hesse (2009) have demonstrated the viability of using LiDAR for detecting archaeological phenomena, particularly in forested environments

where traditional survey methods are limited. Similarly, research by Vinicius-Peripato et al. (2023), and Vaughn & Crawford (2009) has shown that spectral data can provide valuable insights into vegetation patterns that may correspond to past human activity. Given these precedents, this study incorporates both LiDAR and spectral datasets to enhance the detection of architectural and environmental features at Angamuco.

This thesis uses LiDAR data collected in a scan of the Malpais de Zacapu in 2010. The scan was conducted by Merrick Engineering during a single flight and covered a large swath of the region. The airborne LiDAR system, an Optech Titan, was mounted on a plane flying at 900 meters above ground level, covering a 1,000-meter-wide swath. The scan was divided into 500 × 500 meter tiles with 50% overlap to ensure data continuity. The LiDAR pulses were emitted at a 532 nm wavelength (mid-green spectrum) using a discrete return system, achieving a ground sampling rate of 900 Hz. This setup was chosen to balance high-resolution topographic data acquisition while maintaining coverage over Angamuco's rugged terrain.

Alongside LiDAR, I collected mid-resolution multispectral data from LANDSAT 8 Enhanced Thematic Mapper + (ETM+) to supplement the understanding of the canopy structure with spectral signatures. To ensure alignment with the LiDAR dataset, specific selection constraints were applied: the selected raster should have been scanned within a similar timeframe and be minimally obstructed by cloud cover. After evaluating available datasets, the April 16th, 2013, scan from the U.S. Geological Survey (2013) was selected as the best match.

By combining these high-resolution LiDAR and spectral datasets, this study builds on successful research into AASMs in order to engineer the best features for this predictive model.

Label Selection

In machine learning, a label is a designated category or class assigned to a specific data point, representing a known instance of a particular phenomenon. For classification tasks, labels function as the dependent (y) variable that the model attempts to predict. Since this study aims to identify locations of known archaeological typologies, a supervised learning approach is most appropriate; therefore, it requires manually created labels to train the model effectively.

A supervised machine learning model is an approach in which an algorithm learns from labeled data, where each input (X) is associated with a known output (Y). The model iteratively adjusts its internal parameters to minimize the difference between its predictions and the actual labels, thereby improving its ability to classify unseen data. This form of learning is particularly useful when patterns in the data are not explicitly defined but can be inferred through statistical relationships.

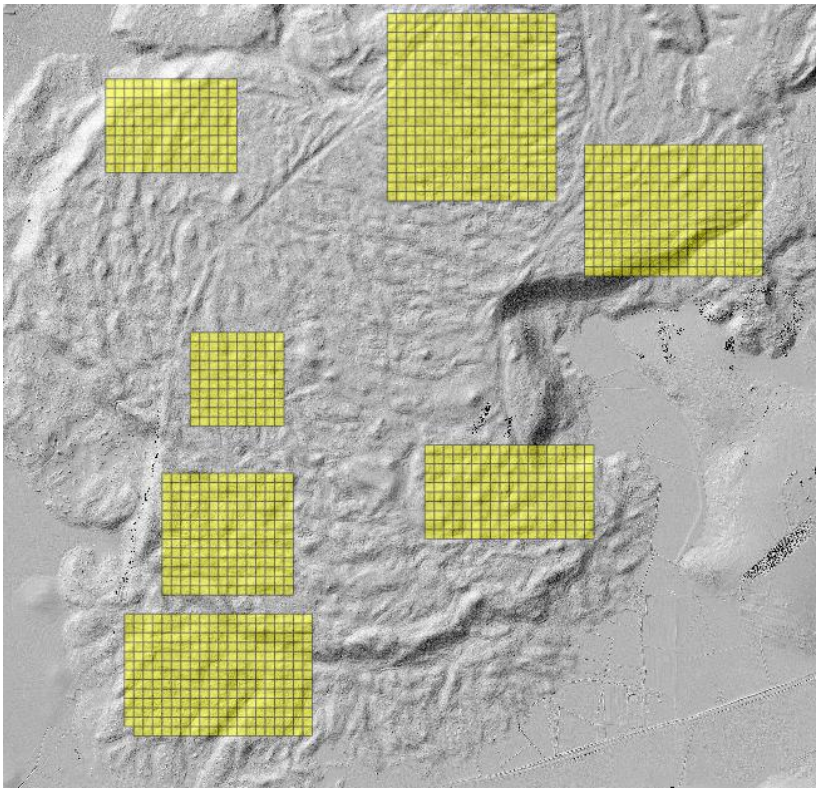


Figure 2: Training areas outlined in yellow. Each cell on the grid was manually labeled using visual analysis. All training areas were chosen in the southern area of the Malpais de Zacapu where extensive research into archtiectural morphology and road network were compiled.

Since this research seeks an indirect method that uses environmental variables as a proxy to detect archaeological features, a per-pixel labeling schema is most appropriate. To implement this, a 30x30m grid-based system was created to emulate

pixels, ensuring that spatial data are organized into discrete, non-overlapping "location bins." Each bin represents a fixed area aligned to the resolution and boundaries of the collected ETM+ raster pixels, maintaining consistency in spatial representation. These bins represent the geographic boundaries for each instance of data—where one cell on the grid is akin to one pixel.

Archaeological phenomena exist as a continuous plane rather than discrete units. Assigning labels based solely on individual cells would risk misrepresenting broader spatial patterns. Instead, label assignment follows a method inspired by convolution, where a cell's label is determined primarily by the attributes of its eight neighboring cells. This approach ensures that labels capture local trends in urban structure rather than being tied to specific cell boundaries, better reflecting the nature of the data. By structuring labels in this manner, the model would theoretically learn more impactful spatial dependencies and relationships between environmental variables and archaeological features.

The dataset is categorized into three primary groups: public space, private space, and terraces. These categories are informed by previous research on the urban design and landscape of Angamuco. The public and private space labels are derived from the work of Solinis-Casparius, who conducted one of the most comprehensive studies of Angamuco's urban organization to date. His research identified areas of public and private significance based on the redundancy of road networks and building typologies, making these categories some of the most obvious and well-supported land-use classifications for labeling (Solinis-Casparius, 2019). This comprehensive approach makes them ideal for predictive modeling, as they provide clear, structured distinctions between urban space types that could be reflected in environmental signals.

The terrace labels are based on the work of Dorison, which highlights the extensive terracing across the site. As much of Angamuco is terraced, these features are a key element of the landscape. Moreover, research by Hightower (2014) has demonstrated that terraces can affect canopy structures, further supporting their inclusion as a distinct category (Dorison, 2022; Hightower et al., 2014).

These categories are not mutually exclusive, meaning that a single area can exhibit characteristics of both public and private spaces or be a mix of urban and terraced landscapes. To better differentiate these complex spatial patterns, each category is divided into subcategories that reduce overlap and highlight meaningful variations in spatial organization. Finally, all categories are binary in order to simplify the task. With 1 (positive) meaning the data instance fits the archaeological land-use typology and a 0 (negative) meaning it does not.

Public Space Subcategories

Public spaces are characterized by redundant road networks, plazas, pyramids, and other ceremonial structures. However, because areas can contain overlapping public, private, and terrace features, public space is divided into subcategories that help distinguish its varying densities and interactions with terraces. These subcategories ensure that labels reflect distinct spatial trends rather than broad, overlapping classifications.

- **Dense Public Space (DensePub):** Cells with 6–8 public neighbors, representing areas with a high concentration of public buildings and/or major redundant roads.
- **Medium Public Space (MedPub):** Cells with 3–5 public neighbors, marking transitional zones between dense public spaces and less developed areas.

- **Sparse Public Space (SparcPub):** Cells with 1–2 public neighbors, indicating peripheral or low-density public areas.

Since terraces have been shown to influence canopy structure (Hightower 2014), public spaces are further divided based on their relationship with terraces:

- **Public (No Terrace) (PubNT):** Public spaces with minimal terrace influence (fewer than three terrace neighbors).
- **Public (With Terrace) (PubWT):** Public spaces interspersed with terraces, representing areas where public infrastructure coexists with terracing.

These subcategories allow the model to distinguish different types of public space and mitigate the overlap between public, private, and terrace classifications.

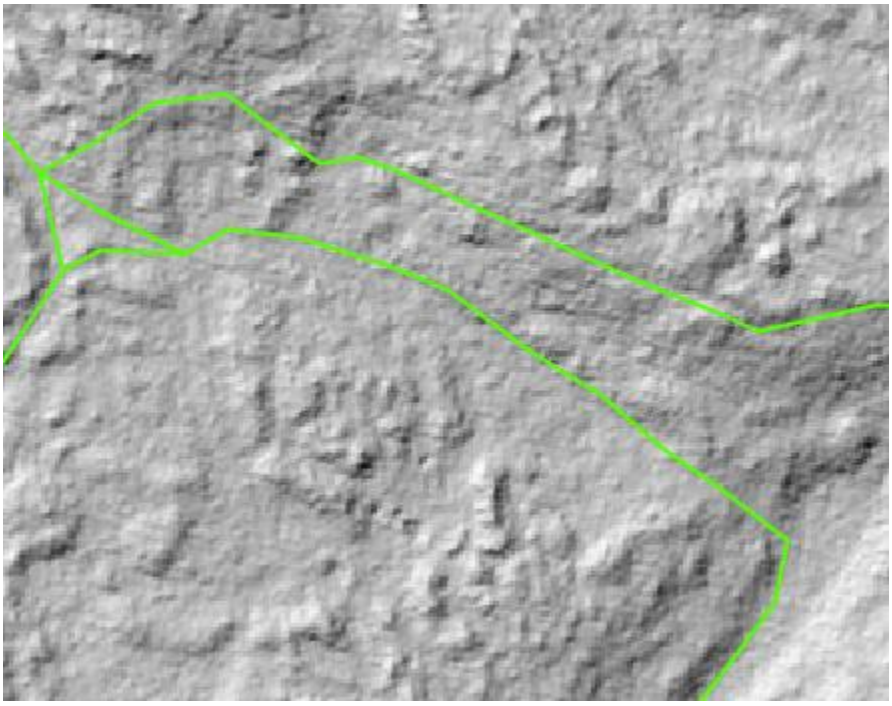


Figure 3: Example of dense public space. This can be described as public space due to the presence of public architecture (plazas) and redundant roads (drawn by Solinis-Casparius 2019).

Private Space Subcategories

Private spaces are defined by areas with less redundant road networks and a high concentration of house mounds. As with public space, private areas are broken down into subcategories that help refine the classification and account for overlap with other spatial features.

- **Dense Private Space (DensePriv):** Cells with 6–8 private neighbors, representing dense residential clusters with minimal major road redundancy.
- **Medium Private Space (MedPriv):** Cells with 3–5 private neighbors, marking transitional zones between high-density private areas and less developed spaces.
- **Sparse Private Space (SparcPriv):** Cells with 1–2 private neighbors, indicating the periphery of residential zones or more loosely clustered households.

To further differentiate private spaces based on their interaction with terraces:

- **Private (No Terrace) (PrivNT):** Private spaces with minimal terrace influence (fewer than three terrace neighbors).
- **Private (With Terrace) (PrivWT):** Private spaces interspersed with terraces, capturing areas where residential structures and terraces coexist.

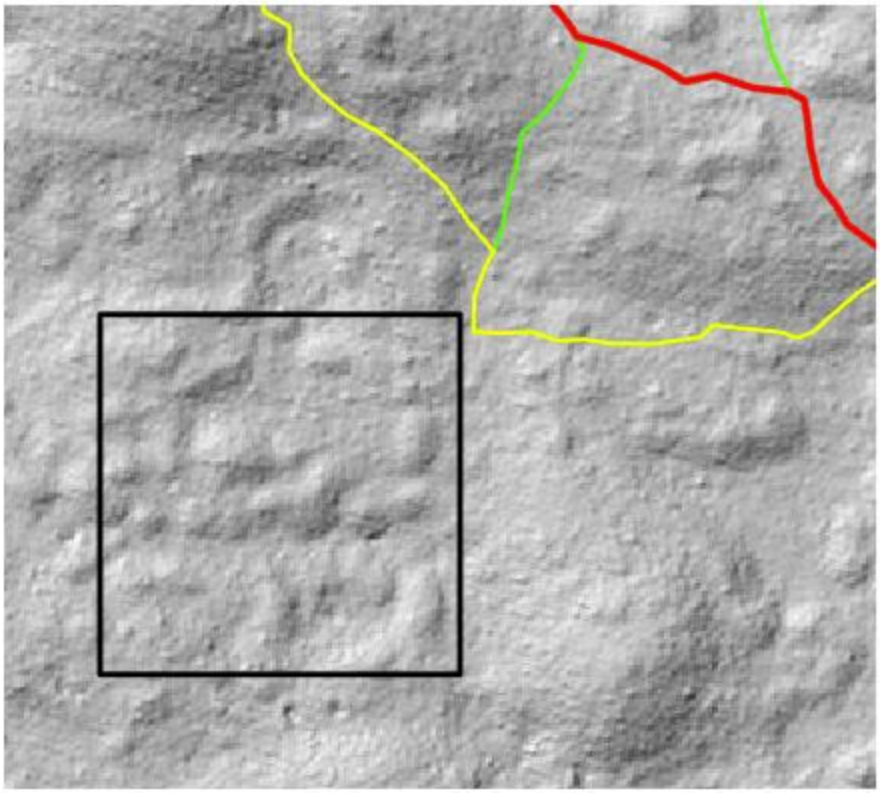


Figure 4: Private space outlined in black. This region is private as it is composed mostly of house mounds, and it is off the path of main (red), secondary (yellow), and tertiary (green) roads. Roads drawn by Solinis-Casparius (2019)

By structuring labels this way, the classification system helps mitigate the inherent overlap between public, private, and terrace spaces, ensuring that each label reflects a distinct spatial pattern while allowing for a nuanced analysis of how these spaces interact.

Terrace Subcategories

Terraces are an important landscape feature at Angamuco, with prior research (Hightower 2014) demonstrating their influence on canopy structure. However, because terraces can be found alongside both public and private spaces, a more detailed classification is necessary to distinguish different terrace contexts. The following subcategories help mitigate overlap and ensure that labels capture meaningful spatial patterns.

- **Dense Terrace (DenseTerr):** Cells with 6–8 terrace neighbors, representing areas with a high concentration of terraces.
- **Medium Terrace (MedTerr):** Cells with 3–5 terrace neighbors, indicating transitional zones between dense terracing and other landscape features.
- **Sparse Terrace (SparceTerr):** Cells with 1–2 terrace neighbors, marking the periphery of terraced areas or isolated terrace features.

Because terraces are present in both urban and non-urban contexts, they are further divided based on their relationship with public and private spaces:

- **Rural Terrace (RuralTerr):** Cells with predominantly terrace neighbors and few urban (public or private) neighbors, representing agricultural or non-urban terracing.
- **Urban Terrace (UrbTerr):** Cells where terraces are interspersed with public or private spaces, capturing areas where terraces coexist with urban infrastructure.

This classification ensures that terrace features are analyzed in relation to both settlement density and their broader spatial context, providing a more refined understanding of their role in the landscape.

Figure 5: Example of terraces. These are a fairly easy category to identify visually due to their easy-to-distinguish morphology.

Labeling Conclusions

This labeling system ensures that spatial classifications reflect meaningful trends rather than rigid, overlapping categories. By structuring public, private, and terrace spaces into distinct subcategories, the model can better capture variations in settlement density and landscape use while accounting for the inherent complexity of Angamuco's urban and agricultural features. These labels provide a foundation for training the model to recognize patterns in canopy structure that correspond to archaeological features.

With the labels defined, the next step is to develop a set of features that can effectively represent the spatial and environmental characteristics of each cell. This process, known as

feature engineering, is crucial for ensuring that the model can identify the relationships between canopy structure and archaeological space.

Feature Engineering

In machine learning, features refer to the input variables (X) used to train a model (not to be confused with archaeological features). Feature engineering is the process of transforming raw data into a structured format that a machine learning algorithm can effectively interpret. For example, an ML algorithm designed to predict locations of archaeological phenomena cannot do this by analyzing raw LiDAR point clouds. Even if it could, without clearly defined features, it would be difficult to analyze the relationships between the input data and the labels being predicted.

Features are constructed in relation to a location bin. Each location bin serves as a single data instance (analogous to a row in a dataset). These bins provide a structured framework to influence how features are calculated. For example, if a feature represents average tree height, its value for a given location bin is determined by averaging all tree heights within that bin's boundaries. This structured system allows for a direct correspondence between spatial data and predictive modeling, enabling the machine learning model to analyze patterns in environmental features and assess the likelihood of archaeological phenomena within each location bin.

Canopy Data Preparation and Feature Engineering

Engineering a concise suite of features to represent a forest remains a complex and challenging process. Before any features can be engineered, an analyst must first create an accurate digital representation of the forest canopy. Since forests vary greatly in terms of tree size and crown structure, any attempt to digitally represent a canopy is context-dependent and requires careful adaptation of existing methods to the specific forest in question.

Preprocessing: Individual Tree Detection

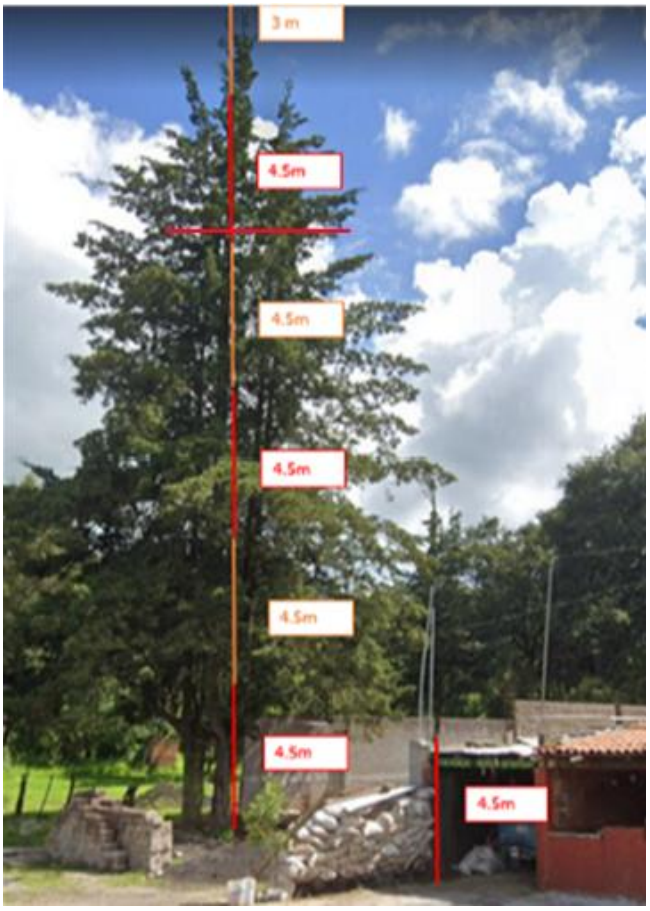


Figure 6: A very crude approximation of tree height and width to calibrate window size. Image taken from Google Earth. The tree and building shown are on a roadside near the edge of the malpais. The 4.5m is derived from average story heights.

The initial step in digitally reconstructing a forest canopy is identifying and segmenting individual trees from the LiDAR point cloud. Individual tree detection and segmentation (ITDS) is inherently complex because it requires domain knowledge about the scanning area to create an accurate simulation. Key factors such as typical tree heights and crown radii are essential for tuning ITDS models. Figure 6 provides a rough estimation of tree height and crown radius, serving as a starting point for the tree detection model. For Angamuco, I estimated a tree height of around 25 meters and a crown radius of approximately 5 meters. I utilized the ITDS framework from the *lidR* library in Python for this process. The performance of ITDS depends on the algorithm used, and based on the data and ecological context of Angamuco, Dalponte (2016) and Silva (2016) offered the most suitable methodologies.

Dalponte (2016) trained a model for ITDS in pine-dominated forests in the Alps to generate a carbon budget. He developed a canopy height model by classifying ground and

canopy points from a LiDAR point cloud, followed by creating a raster of the classified canopy points. Dalponte's method uses Local Maxima Filtering (LMF) to identify the highest points within a user-defined window size. The algorithm segments trees by comparing the elevation of points to their neighbors, starting from the local maximum. As the algorithm radiates outward, points that are lower than their neighbors are classified as part of the same tree crown. When the elevation of neighboring points begins to rise, the algorithm classifies them as belonging to a separate tree (Dalponte & Coomes, 2016).

Silva (2016) applied a similar approach to Dalponte's, using LMF to segment trees in southwestern Georgia. However, Silva incorporated machine learning (ML) to enhance crown delineation. Unlike Dalponte's method, which relies solely on LMF and predefined rules, Silva used a random forest algorithm to assess the relationship between LiDAR points and their local maxima. This machine learning approach allows Silva's method to adaptively classify points as part of a tree crown, improving the segmentation of more irregular or variable crown structures (Silva et al., 2016). While Dalponte's method assumes a consistent radial structure for tree

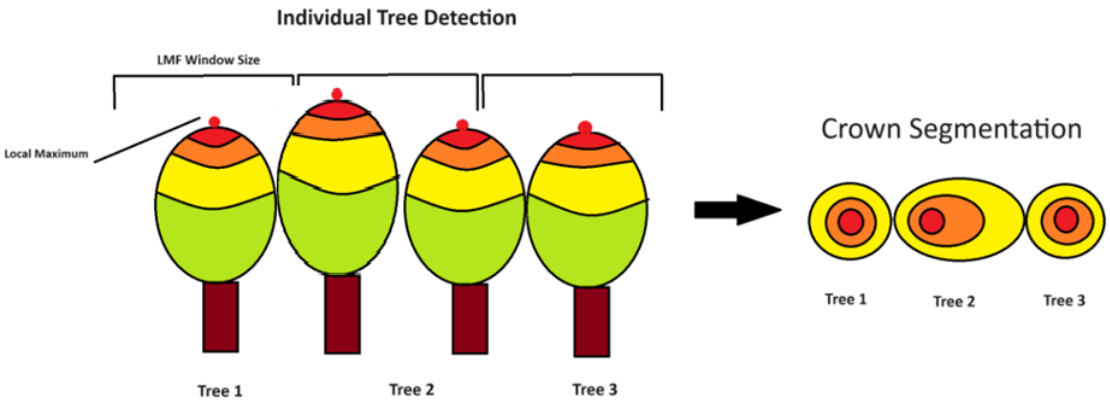


Figure 7: A demonstration of the general workings of an IDTS algorithm a poorly calibrated LMF window size. This model fails to recognize one of the trees and counts the middle two trees as a single tree. Made in MS paint.

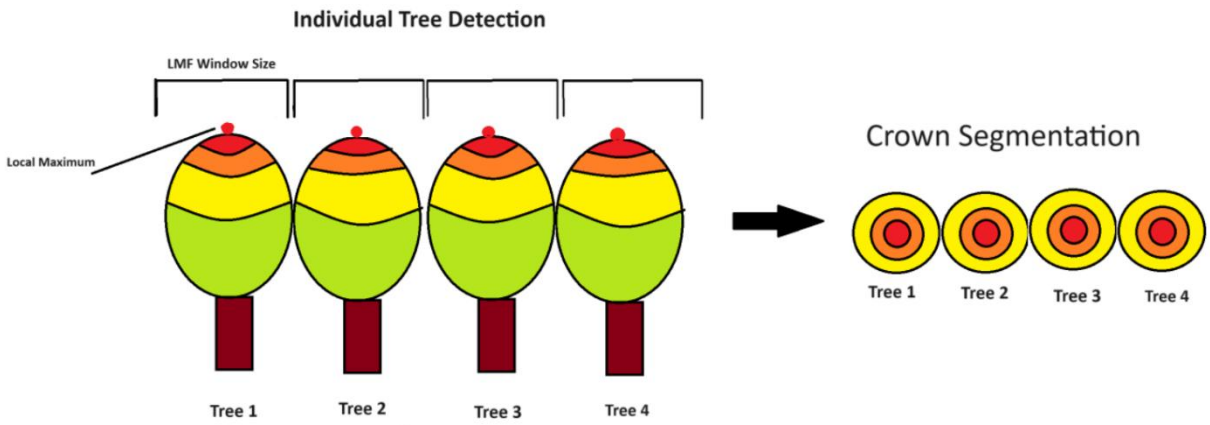


Figure 8: A demonstration of the general workings of an LMF IDTS algorithm with an ideal window size. This model perfectly captures all trees in the dataset.

crowns, Silva's model, through ML, accounts for more variability in crown shape, making it better suited for forests with diverse tree structures.

The choice of LMF window size is critical, as demonstrated in Figures 7 and 8. If the window is too large, two trees may be erroneously combined, and the taller crown may be selected as the local maximum. Conversely, a window that is too small may identify multiple local maxima on a single tree, leading to overcounting. This misstep in tree detection can then negatively affect crown segmentation. Both Silva and Dalponte validated their methods through ground-truthing, confirming their accuracy. However, ground-truthing was beyond the scope of this research. Given that these algorithms have been ground-truthed in previous studies, visual analysis of remote sensing data offers a reliable alternative for assessing the accuracy of the ITDS model in Angamuco. Figure 9 shows the area of interest,

Figure 9: Google Earth image of the chosen area on the Malpais de Zacapu for visual analysis of the ITDS.

which contains a mix of treed and treeless regions, along with trees of various heights. These factors make visual analysis effective in detecting errors in tree identification and crown delineation.

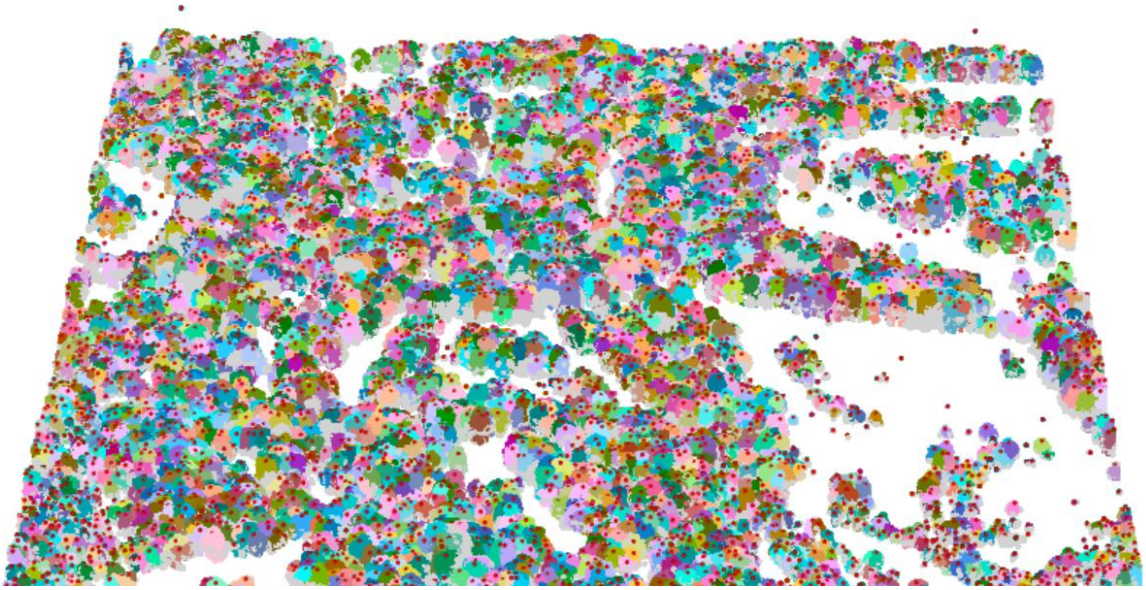


Figure 10: Segmented LiDAR point cloud with classified tree IDs of the visual validation area based on dalponte2016 algorithm with a LMF window size of 5.

Method 1: My initial attempt at ITDS for Angamuco utilized Dalponte's algorithm with a static LMF (Figure 6). Visual analysis confirmed that Method 1 successfully captured the uniquely shaped voids in the canopy and identified smaller trees within those voids. Figure 10 presents a LiDAR point cloud processed by the Dalponte (2016) model with an LMF of 5, segmenting individual trees. Some areas appear "greyed out," indicating regions where the model could not segment trees. These gaps could significantly impact subsequent feature engineering efforts.

Method 2: Despite method 1's reasonable success, I proceeded with a second method to better classify pixels. For method 2, I employed Silva's algorithm with a dynamic LMF (figure 7) to account for crown size variability based on tree height. The dynamic LMF is defined by the function:

$$f(x) = (x * 0.1) + 3$$

where x represents tree height and f(x)defines the LMF window size.

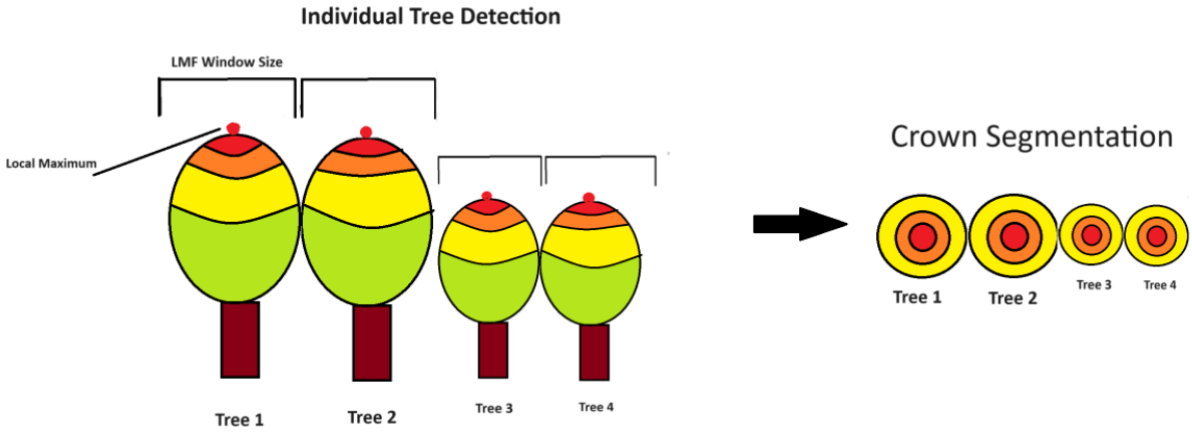


Figure 11: Demonstration of an IDTS algorithm with a variable window size. This model accounts for trees of different sizes.

Figure 11 illustrates the logic of a dynamic window size. This dynamic size along with Silva's algorithm appeared to produce better results. Figure 12 shows the LiDAR point cloud processed by Method 2. This figure zooms in on an area where Method 1 struggled to segment trees, showing fewer "greyed-out" regions, thus indicating that Method 2 performs better in the conditions at Angamuco, as more of the LiDAR points are classified to specific trees.

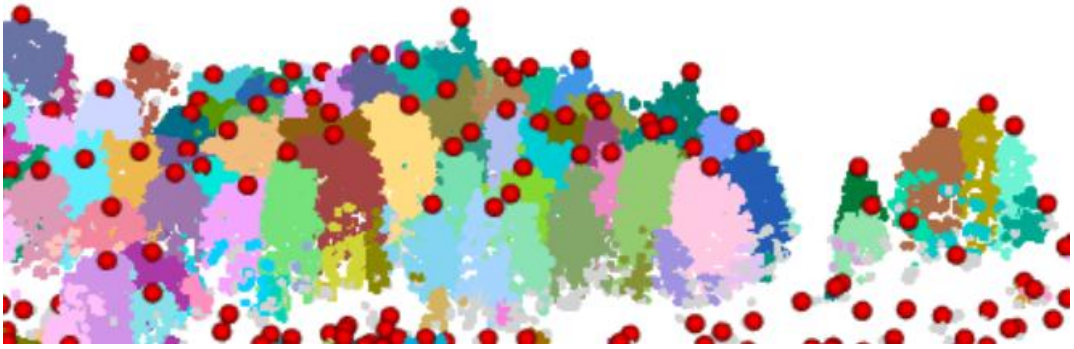


*Figure 12: Digital Canopy Model with classified tree IDs from a planar based on silva2016 algorithm with an lmf of f(x) = {x * 0.1 + 3}. The red dots represent the local maximum used to identify individual trees.*

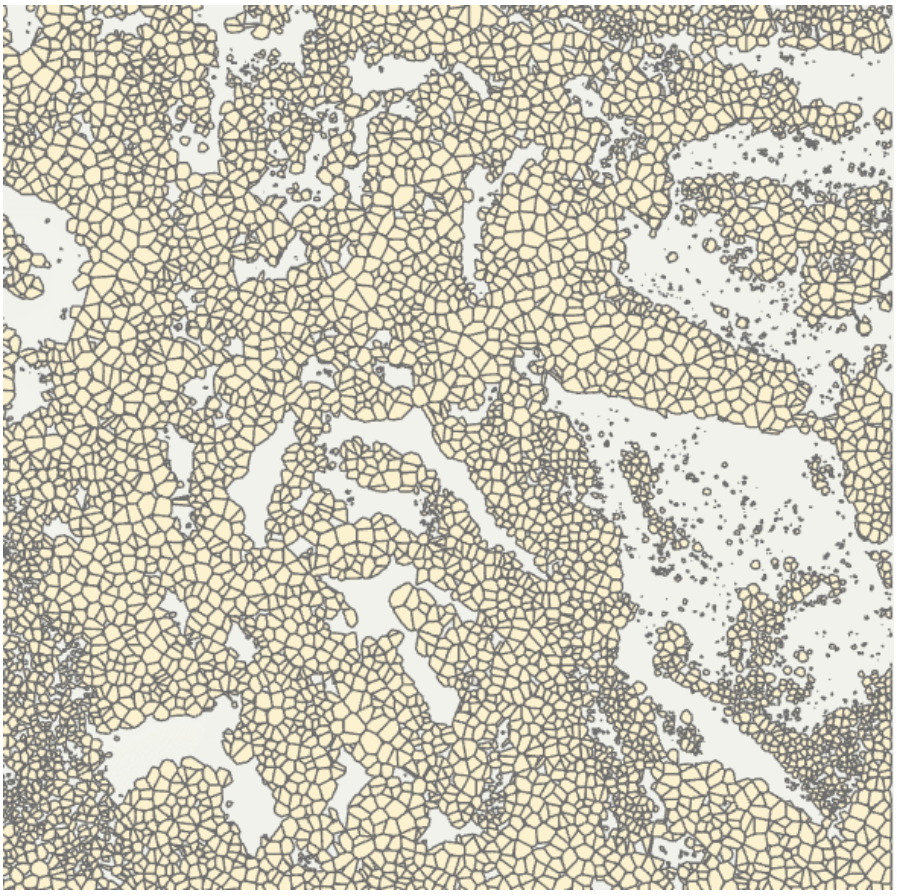


Figure 13: Crown delineation shapefile using method 2

Finally, Figure 13 shows the tree crown boundaries delineated by Method 2. The shapefile reveals the diversity in crown shapes and sizes, a key quality for the proposed AASM at Angamuco, which relies on crown variability as a defining feature.

Canopy Feature Engineering

Method 2 of the IDTS preprocessed the raw LiDAR data to create a usable digital canopy model with segmented trees and delineated crowns. The next step is to calculate features from this model. The data within each bin needs to be transformed into a suite of concise features that represent the canopy structure. Moran et al. (2018) devised six features that sufficiently describe forest canopies for modeling purposes. Moran sought to describe the Horizontal and Vertical aspects of the canopy

- Vertical features refer to the distribution of tree heights within a location bin, representing the canopy's structure along the z-axis (the vertical dimension). These features capture how tree heights vary within a given area, reflecting the vertical stratification of the canopy.
- Horizontal features, on the other hand, describe the distribution of canopy structure along the x and y planes of the landscape. These features capture how the density and spatial arrangement of tree crowns vary within the horizontal plane, offering insight into the spread and uniformity of the canopy at ground level.

They aimed to reduce the dimensionality of the data to avoid overlap and reduce a “black-box” effect. Moran settled on four features to describe the vertical structure of the canopy and two to describe the horizontal. The four vertical features use L-Moments (L1, L2, T3, T4) to describe the distribution of values across the vertical dimension. The horizontal features include the

average density of LiDAR points and the standard deviation of LiDAR density to capture the horizontal components of canopy structure. These horizontal features help establish a baseline density and capture variability within the location (Moran et al., 2018).

L-moments are a series of four concise descriptive statistics:

- **L-Location (L1)**: The mean value of the dataset, which in our context represents the mean height of all the trees in a location bin.
- **L-Scale (L2)**: Measures the dispersion of the dataset, indicating whether values change gradually or abruptly. In our context, this feature describes whether trees in a location bin have a consistent height gradient or exhibit abrupt height differences.
- **L-Skew (T3)**: Measures the skewness of the data, showing if trees tend to be higher or lower than the mean height in a location bin.
- **L-Kurtosis (T4)**: Measures the kurtosis or 'weight of tails' in the distribution. This indicates how many trees are around the mean, specifically whether tree heights tend to be clustered near the extremes (Hosking, 1990).

I adapted Moran's single horizontal density feature into two: 1) the average number of LiDAR points per square meter in each location bin and 2) the standard deviation of that distribution. These two features provide insight into how consistently dense the canopy is across a location bin. For percent canopy cover, I divided the area of tree crowns (as generated by the IDTS) by the area of the location bin. Additionally, I included a feature to report the number of trees within each bin based on the tree top points derived from the LMF.

Finally, I incorporated open-source spectral data. Given that previous research indicated a reasonably homogeneous forest canopy at Angamuco, trees growing on archaeological features may reflect distinct spectral signatures due to variations in growing conditions. Since the pixels in the multispectral ETM+ raster aligned with the locations bins, I simply associated the reflectance values from six bands (blue, green, red, NIR, SWIR 1, and SWIR 2) with each location bin, then combined these reflectance values into a single "spectral signature" using principal component analysis (PCA).

The process described in this section resulted in nine features to represent the canopy structure at the 30m x 30m scale.

Topographic Features

While canopy structure features are the main focus of this research, intuition suggests that features describing topographic changes over space could help identify areas where architectural archaeological phenomena may be present. The process of engineering topographic features from LiDAR data is relatively straightforward. Raw LiDAR data generates a point cloud representing all physical features of a scan area: trees, underbrush, mountains, hills, buildings, etc. Points not reflected off the ground must be filtered out to derive the topographic dynamics of the surface. Various classification protocols exist to classify ground points, typically focusing on the last laser pulse, which returns to the sensor as if reflecting off ground locations. For this research, I used ESRI's ground classification algorithm. Once ground points remain, the resulting elevation data can be transformed into a Digital Elevation Model (DEM), which is then used to engineer topographic features.

Although these topographic features are secondary to the more detailed features describing canopy structure, they could offer useful insights to narrow down macro-level decisions about terrain types that could indicate specific archaeological phenomena. I created three features based on topographic considerations. First, the range of elevation values provides a measure of how flat or uneven a location bin is, helping identify areas of flat terrain or rough, variable terrain. Second, the standard deviation of elevation values captures how gradually the elevation changes across a location, distinguishing between gentle slopes and abrupt changes. Lastly, an interaction feature constructed by multiplying the elevation range by the elevation standard deviation to differentiate between rough terrain with flat areas and rough terrain with more gradually sloped areas, potentially narrowing the search for specific types of archaeological sites.

ANOVA Analysis

To validate the existence of statistical relationships between the engineered features and the assigned labels, I conducted a series of one-way ANOVA (Analysis of Variance). ANOVA is a statistical method used to determine whether there are significant differences in the mean values of continuous variables across multiple categorical groups. Performance of an ANOVA test is typically measured through *p,* where a low p-value (typically <0.05) indicates that at least one category has a mean that differs significantly from the others. ANOVA is particularly useful in this context to assess if related labels are meaningfully different from one another based on the variance of each of the 12 features.

As ANOVA tests assume normality, I transformed some of the features so that the distributions are more normal. Figure 14 reports the list of 12 features that will be used for the ANOVA (and later predictive modeling). This table reports a shortened codename, a brief description along with any transformation, and the general type of feature.

Codename	**Description**	**Feature type**
L1	L-Location	Lidar/ Vertical Canopy
L2	L-Scale	Lidar /Vertical Canopy
T3	L-Skew	Lidar /Vertical Canopy
T4	L-Kurtosis	Lidar /Vertical Canopy
AvgCrown_density	Average Crown Density	Lidar/ Horizontal Canopy
stdCrown_density	σ of Average Crown Density	Lidar/ Horizontal Canopy
BOXcov_percentage	Box-Cox transformed % Canopy Cover	Lidar/ Horizontal Canopy
TreeCount	# of Trees in the Location Bin	Lidar/ Horizontal Canopy
PCA_Bands	PCA of 6 Spectral Signatures of a Location Bin	Spectral / Canopy
LOGelev_range	Log-transformed elevation range	Lidar/ Topographic
LOGelev_std	Log-transformed elevation σ	Lidar/ Topographic
LOGERxS	Log-transformed interaction (range* σ)	Lidar/ Topographic

Figure 14: The 12 environmental features

For each ANOVA test, categorical labels were grouped into factor variables based on shared characteristics. The six ANOVA procedures were as follows: (1) Terrace Density (sparseTerr, medTerr, denseTerr, na), (2) Terrace Urbanism (ruralTerr, urbTerr, na), (3) Private Density (sparsePriv, medPriv, densePriv, na), (4) Terraced Private (privNT, privWT, na), (5) Public Density (sparsePub, medPub, densePub, na), and (6) Terraced Public (pubNT, pubWT, na). The dependent variables were the 12 engineered features.

A statistically significant result in ANOVA is determined by a low p-value (typically <0.05), indicating that at least one category has a mean that differs significantly from the others. The F-statistic, which represents the ratio of variance between groups to variance within groups, provides a measure of how strongly the categorical grouping explains variance in the feature of interest.

Across all tests, PCA_bands consistently had the highest F-statistic and lowest p-value, suggesting a strong relationship between spectral variation and label categories. Additional features, including Boxcov_per, AvgCrown_d, L1, and L2, were also frequently significant.

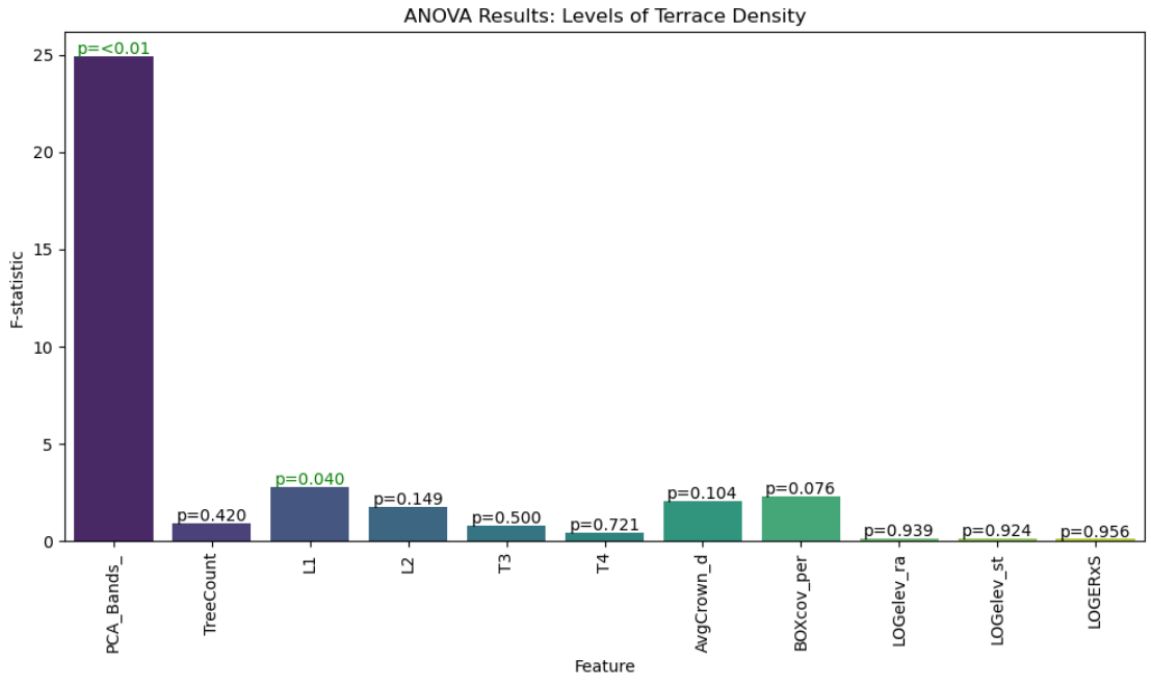


Figure 15: Terrace density ANOVA

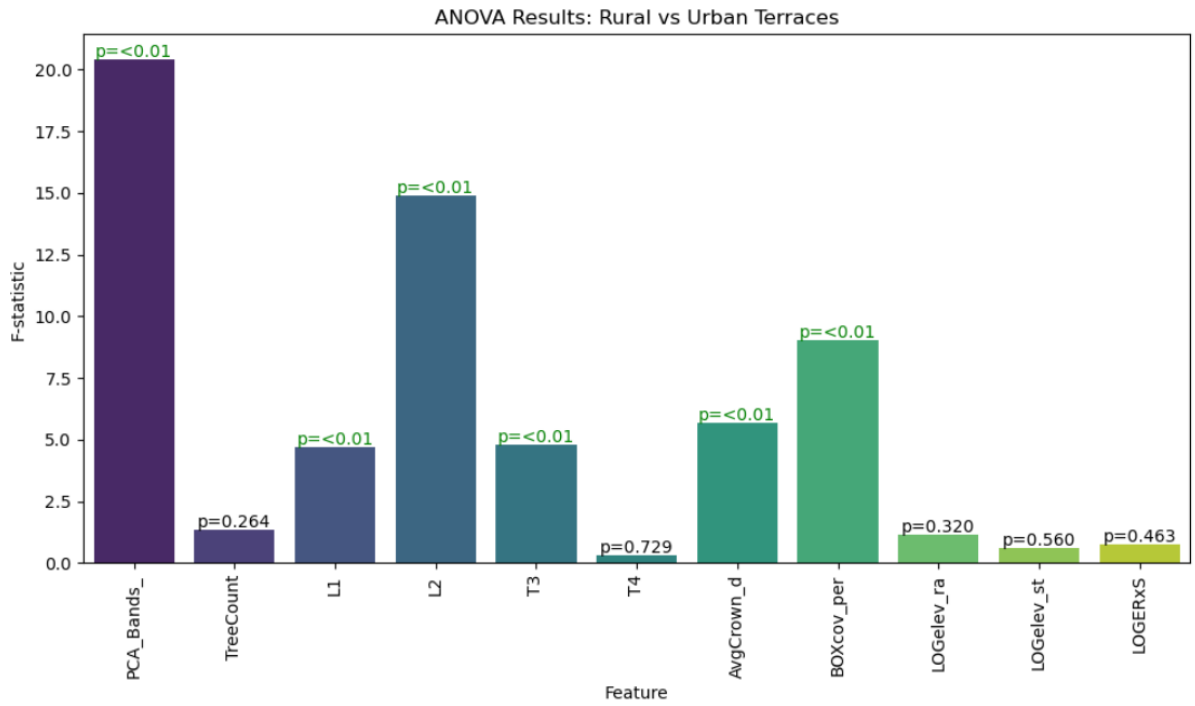


Figure 16: Rural vs urban terrace ANOVA

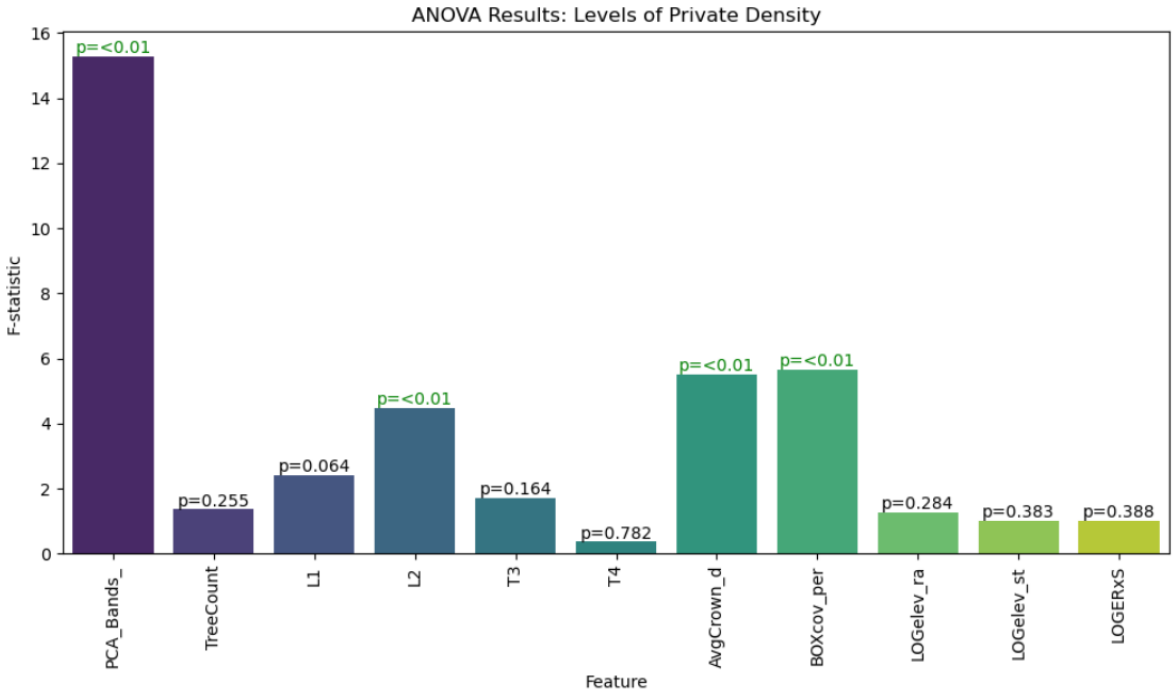


Figure 17: Private space density ANOVA

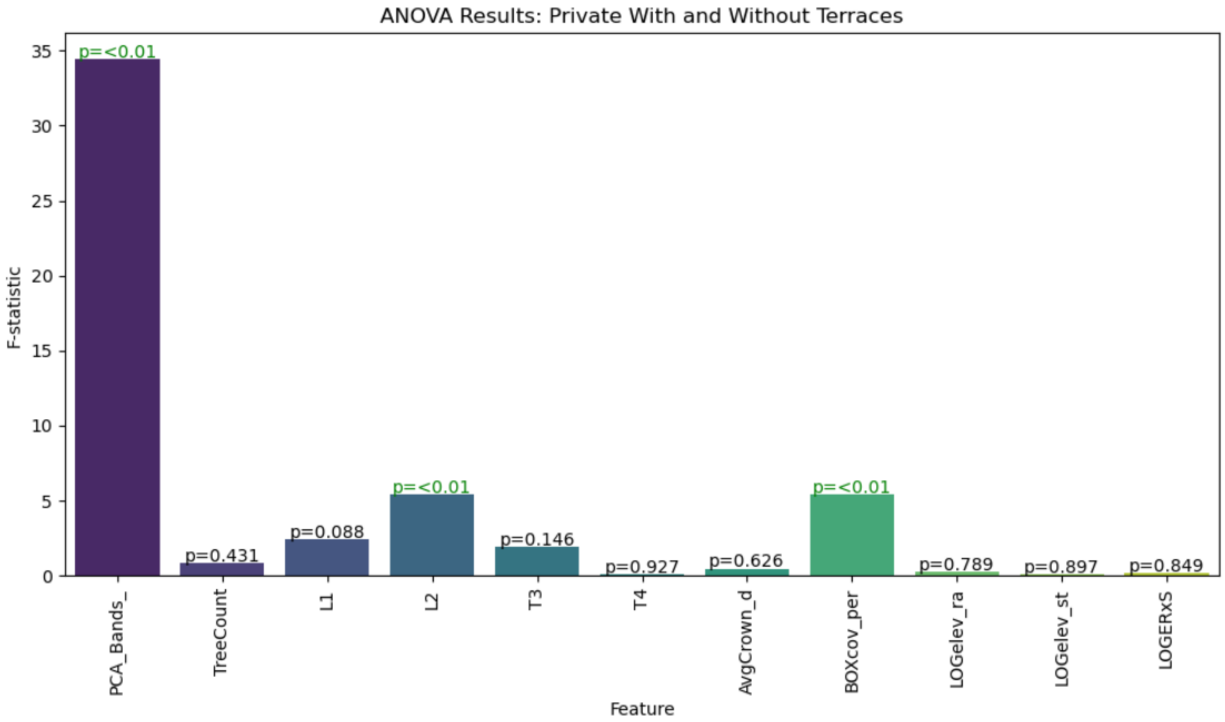


Figure 18: Private space with and without terraces ANOVA

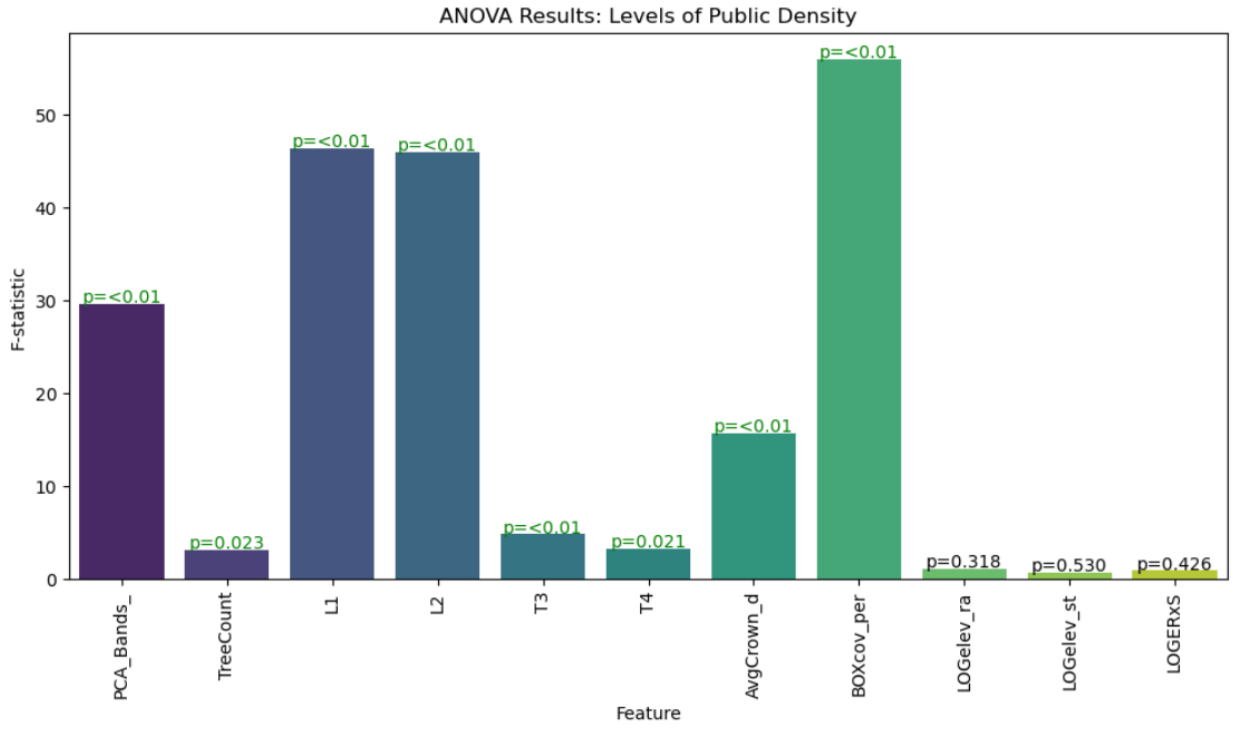


Figure 19: Levels of public space density ANOVA

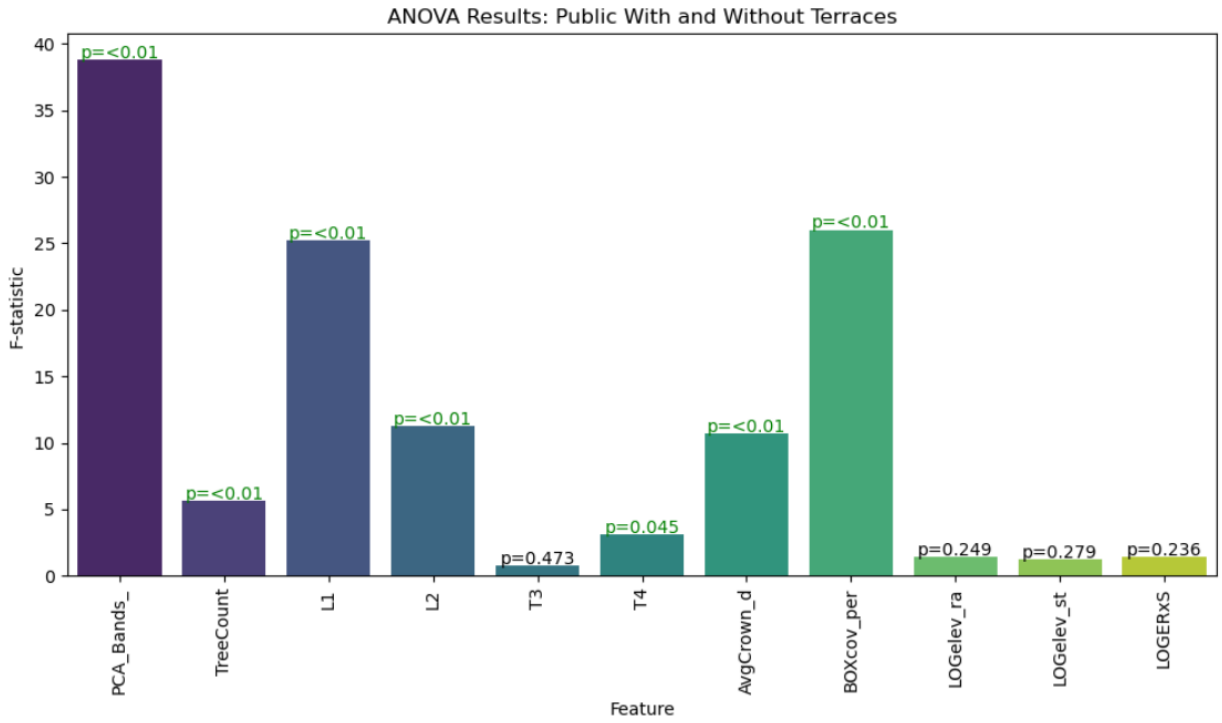


Figure 20: Public Spaces with and without terrace ANOVA

Interestingly, topographic features were never found to be statistically significant. While this suggests that topography alone may not differentiate between label categories, it does not imply that these features lack predictive value. ANOVA assumes linear relationships and evaluates features in isolation, meaning that interactions between variables are not accounted for. In a predictive modeling context, topographic variables may still contribute valuable information when used alongside canopy-based features. Therefore, this ANOVA analysis serves as a preliminary validation that relationships exist between the engineered features and the labels, supporting their use in subsequent predictive modeling.

PerMANOVA Analysis

I additionally performed a Permutational Multivariate Analysis of Variance (PerMANOVA) to confirm which individual labels held the most significant relationships to the

suite of 12 features. This statistical approach helps to identify which labels exhibit the strongest and most consistent patterns, narrowing down the most relevant labels for predictive modeling. By quantifying the relationship between environmental features and archaeological typologies, the analysis provides a solid foundation for selecting the most effective labels for the model's predictive tasks.

This approach mirrors Hightower's (2014) analysis, which linked terraces to canopy structures. PerMANOVA is a non-parametric statistical test that compares group differences based on a distance matrix. Unlike traditional ANOVA, which assumes normality and equal variances, PerMANOVA offers more flexibility for analyzing ecological and spatial datasets. Researchers frequently use this method to examine relationships between environmental variables and categorical groupings in fields like ecology, landscape archaeology, and remote sensing (Anderson, 2001).

Label	Sum of Squares	R-Squared	F-Statistic	P-Value
densePub	320.900	0.029	44.985	< 0.001
medPub	33.800	0.003	4.618	0.009
sparPub	124.700	0.011	17.167	< 0.001
pubNT	128.800	0.012	17.736	0.001
pubWT	1.600	< 0.001	0.212	0.890
densePriv	11.700	0.001	1.599	0.185
medPriv	9.200	< 0.001	1.249	0.304
sparPriv	14.000	0.001	1.912	0.136
privNT	5.500	< 0.001	0.752	0.504
privWT	9.600	< 0.001	1.303	0.270
denseTerr	2.400	< 0.001	0.323	0.800
medTerr	4.200	< 0.001	0.578	0.629
sparTerr	8.700	< 0.001	1.190	0.284
ruralTerr	45.400	0.003	4.789	< 0.001
urbanTerr	14.100	0.001	1.482	0.171

Figure 21: PerMANOVA results for individual labels against the suite of 12 environmental features.

By running PerMANOVA on the labeled dataset, I identified the labels that explain meaningful variance in canopy structure. This approach ensured that I focused the model on labels that capture significant spatial trends while removing those with weaker associations. Labels with high pseudo-F values and R^2 scores proved most relevant, while those with low explanatory power were either deprioritized or merged with other categories.

The results of the PerMANOVA analysis highlighted several key insights regarding the spatial labels in my dataset. For the public space category, the densePub label exhibited the strongest relationship with variations in canopy structure, as indicated by the lowest p-value and

highest pseudo-F-statistic. This suggests that areas characterized by dense public space, with redundant road networks and significant ceremonial structures such as plazas and pyramids, are the most distinct in terms of their influence on canopy features. The dense public spaces appear to create a more robust pattern in canopy structure compared to other public space subcategories, making densePub a critical label for model training in this context.

In the terrace category, the ruralTerr label was the only significant one. This finding suggests that rural terrace areas, with predominantly terrace neighbors and few urban features, have a meaningful effect on the canopy structure, or ruralTerr is the most well-defined terrace label in its relationship to canopy structure. Interestingly, no private space labels were found to be significant, indicating that private areas, which typically feature less redundant road infrastructure and more residential structures, do not have as strong a correlation with canopy variation in this case.

Despite the lack of significant private space labels, simple intuition—supported by initial algorithmic tests, including gradient boosting—pointed to PrivNT (Private Space with No Terrace) as a promising label for further analysis. This label may still offer valuable insights when used in conjunction with other labels in predictive models.

Algorithm Selection

The perMANOVA analysis demonstrated a statistically significant relationship between some labels and the suite of features. However, the low R-squared values indicated that these relationships were highly complex and nonlinear, suggesting that traditional linear algorithms would be insufficient. Instead, an algorithm capable of handling multiplex, nonlinear relationships was necessary. This conclusion aligns with insights from previous research, which

suggested that the interactions between topographic and vegetation features at Angamuco are varied and multifaceted. The following section discusses the models tested in this study, their theoretical suitability, and their observed performance.

Random Forest

Random forests are ensembles of decision trees designed to improve predictive accuracy by leveraging multiple model outputs. Each tree within a random forest differs slightly in structure due to randomization in the training process. The ensemble nature of random forests helps address three key challenges in modeling complex data (Dietterich, 2002):

1. When multiple different outputs yield similar model accuracies.
2. When searching for the best possible solution is computationally prohibitive.
3. When the true underlying relationship is too intricate for a single model to reconstruct.

By combining multiple decision trees, random forests simulate the complexity of true data relationships while maintaining computational efficiency. Each data vector runs through all trees, and the final prediction is determined by a majority vote or an averaged output (Biau & Scornet, 2016). Given the complex relationships in the Angamuco dataset, a multiplex predictor like a random forest theoretically appeared well-suited to the problem.

However, the random forest ultimately failed to capture meaningful relationships between features and labels at Angamuco. Like decision trees, random forests are deterministic classifiers, meaning they lack the flexibility of probabilistic models. Additionally, the inability of decision trees and random forests to iteratively adjust and learn from data nuances limited their effectiveness. These shortcomings suggested that a non-iterative model might not be capable of accurately modeling the intricate and highly specific relationships within the dataset.

Gradient Boosting

Gradient boosting improves upon the decision tree framework by leveraging an iterative ensemble approach based on boosting theory (Schapire, 2003). Unlike random forests, which construct independent trees, gradient boosting builds successive models that correct the errors of previous models. This iterative process enables the model to learn increasingly complex relationships within the data. The key advantage of boosting is its ability to create a robust predictive framework by combining multiple "weak" learners into a strong overall prediction. The key difference in boosting algorithms compared to other iterative models is that it seeks models that work well in an ensemble instead of models that work well alone. This feature helps the algorithm in preventing overfitted models. Figure 19 outlines this logic.

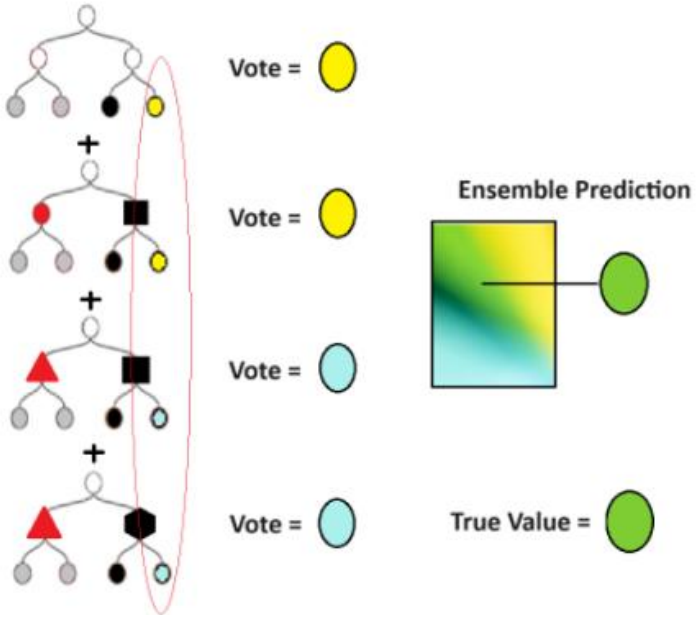


Figure 22: Demonstration of Gradient Boost. Here the learned mathematical relationships are represented as shapes and colors. Rather than trying to vote green, the algorithm seeks models that will vote yellow and blue to create an ensemble vote of green. Made in MS paint.

Gradient boosting was well-suited to the Angamuco dataset for several reasons. First, the iterative nature of the model allowed it to adapt to nuanced relationships that were otherwise overlooked by deterministic classifiers. Additionally, the implementation used in this study, employs a probability score (an output between 0-1 that represents a predicted probability that an instance is either positive or negative) approach, allowing for finer distinctions between classes.

Exploratory analysis revealed that the gradient boost model was particularly effective in predicting negative instances, although its sensitivity to positive predictions remained an issue.

Neural Network

Neural networks optimize data relationships through gradient descent, backpropagation, and the chain rule. These processes enable neural networks to iteratively adjust variable weights in response to errors, improving prediction accuracy over successive epochs. The fundamental advantage of neural networks lies in their ability to model highly specific, nonlinear relationships between features and labels (Lecun, 2001). Similar to gradient boosts, neural networks can produce a probability score for binary classification to create more nuanced predictions. Unlike gradient boosting, which constructs an ensemble of simple models, a neural network seeks to build a single, highly complex model that minimizes predictive loss.

In the context of Angamuco, neural networks demonstrated some potential for predictive modeling, outperforming random guessing. However, they underperformed compared to gradient boosting, particularly in predicting negative instances. Additionally, neural networks are prone to overfitting, as they may learn highly specific patterns that do not generalize well beyond the training data. Despite this limitation, neural networks do have the capability of fitting some of the most complex models, which might be the case given the low R^2 scores from the PerMANOVA.

Model Directions

Given the PerMANOVA results and the need for a model capable of handling multiplex, nonlinear relationships, gradient boosting and neural networks emerged as the most suitable algorithms for archaeological feature prediction at Angamuco. They both feature an iterative

ensemble approach which proved to be the best framework for modeling the intricate and varied interactions between topographic and vegetation features in the study area.

Ultimately, this exploratory analysis reaffirmed the complexity of the task. Only the gradient boost and neural network algorithms demonstrated the potential for generating predictions better than random guessing. However, neither model achieved a stable balance—both exhibited a tendency to overpredict either positive or negative instances. Attempts to fine-tune hyperparameters exacerbated this instability, underscoring the challenge of finding an optimal configuration.

A key insight emerged from the gradient boost model's performance: while it consistently and accurately identified negative instances, it struggled to correctly predict positive instances. This asymmetry suggested a possible solution—rather than relying on a single fitted model, a multi-step approach could be more effective.

The proposed solution was to introduce a second fitted model into the process. The first gradient boost model would act as a coarse filter, reliably excluding negative cases. The second model would then focus on distinguishing between the positive predictions made by the first model, learning to differentiate between true positives and false positives. By training on this more nuanced subset, a second model could refine predictions and mitigate the sensitivity issues inherent in archaeological site modeling. Given the complexity and subtlety of relationships within this refined dataset, a neural network was identified as the most suitable choice for the second stage of classification.

This insight led to the development of an innovative multi-modal architecture: a sequential machine learning pipeline that begins with a Gradient Boost model and concludes

with a Neural Network. This approach leverages the strengths of both algorithms—the robustness of ensemble learning in the first stage and the adaptability of deep learning in the second—to capture the multiplex, nonlinear relationships between environmental signals and archaeological features at Angamuco.

Evaluation Methods

Statistical Evaluation

The statistical evaluation of the model's performance played a crucial role in iteratively improving both the model architecture and the selection of hyperparameters (settings that configure the algorithm before training a model which influences how well that model can learn and make predictions).

To address the challenges of class imbalance—where one label class significantly outnumbers the other—I selected a suite of evaluation statistics designed to balance recall and precision.

- **Confusion Matrix**: Though not a test in itself, the confusion matrix is a vital tool for visualizing prediction outcomes. It shows how many predictions fall into true positives (predicted pos, actual pos); true negatives (predicted neg, actual neg); false positives (predicted pos, actual neg); and false negatives (predicted neg, actual pos), helping to assess where the model is making errors.
- **MCC (Matthews Correlation Coefficient)**: MCC evaluates the quality of binary (two-class) classifications. It ranges from -1 (indicating a completely incorrect prediction) to +1 (indicating perfect predictions), with 0 representing random guesses.

- **Kappa**: This statistic measures agreement between predicted and actual classifications while accounting for chance. A Kappa value close to 1 indicates strong agreement, while values closer to 0 suggest agreement no better than random chance.
- **F1 Score**: The F1 score is the harmonic mean of precision and recall, balancing the trade-off between the two. A high F1 score indicates that both false positives and false negatives are minimized, which is important in situations of class imbalance.
- **AUC (Area Under the Curve)**: AUC evaluates the model's ability to distinguish between classes across various prediction thresholds. The threshold, mathematically represented by θ, is the value between 0 and 1 at which the model assigns a positive or negative label from the output prediction score; adjusting this threshold can affect precision and recall. AUC ranges from 0 to 1, with higher values indicating better model performance. AUC is particularly useful in class-imbalanced scenarios as it evaluates model performance across all thresholds, making it more robust to varying class distributions. Essentially AUC evaluates how likely a model associates higher prediction scores with true positive data instances and vice versa. An AUC score of 0.8 means that there is an 80% chance that a randomly selected true positive will be ranked higher in prediction score than a randomly selected true negative. In other words, the model is correct 80% of the time when distinguishing between the positive and negative classes."

Among these metrics, AUC was the most preferred as it handles class imbalance well and provides a comprehensive evaluation of model performance across different prediction thresholds. This flexibility allows it to capture a model's overall ability to discriminate between positive and negative classes.

Method for Visual Evaluation

To further assess the performance of the BPE model on classification, I conducted a visual evaluation of its predictions on unlabeled data. I selected 200 positive predictions (instances where the model predicted the presence of some archaeological phenomena as defined by the label) and 200 negative predictions (location bins where the model predicted the assigned label did not exist). Additionally, I separated the study area into two regions: the north region and the south region, divided at a latitude of 2.169e6. The south region contained all the training data, whereas the north region had no training data, allowing for an assessment of how well the model generalizes to unseen areas within the typologically varied Malpais de Zacapu.

For the selection of predictions, I used a stratified random sampling approach. In the north region, I randomly selected 200 positive predictions and 200 negative predictions to ensure a balanced representation. I then conducted another stratified random sample for the south region, maintaining the same balance between positive and negative predictions.

For the positive predictions, I evaluated each instance based on three criteria:

1. Correct vs. Incorrect Prediction: I manually determined whether the model correctly identified the assigned label.

2. Almost Correct Prediction: As most labels are not mutually exclusive, an almost correct prediction was a way to recognize a prediction that is not an exact match to a given label but still closely aligns within related categories. For instance, some related labels may account for a large number of “false positive” predictions. This criterion acknowledges the inherent complexity in archaeological classification, where boundaries between categories may be fluid or ambiguous. Thus, an "almost correct" prediction is recognized

as contributing meaningfully to the overall model accuracy, even if it doesn't perfectly align with the predefined label.

3. Identification of Any Archaeological Phenomenon: I merely assessed whether a positive prediction identified some form of archaeological feature (1) or failed to identify any meaningful archaeological phenomenon (0).

For the negative predictions, I applied a single evaluation criterion:

- Correct vs. Incorrect Prediction: I determined whether the model's negative classification was accurate (1 for correct, 0 for incorrect).

By incorporating regional separation and stratified random sampling, this evaluation not only provides insight into the model's accuracy but also examines its ability to generalize across different geographic areas. This distinction is particularly important for assessing whether the model has overfitted to patterns in the training region or if it can successfully identify labels in novel, unseen locations.

Bi-Path Ensemble (BPE) Model Architecture

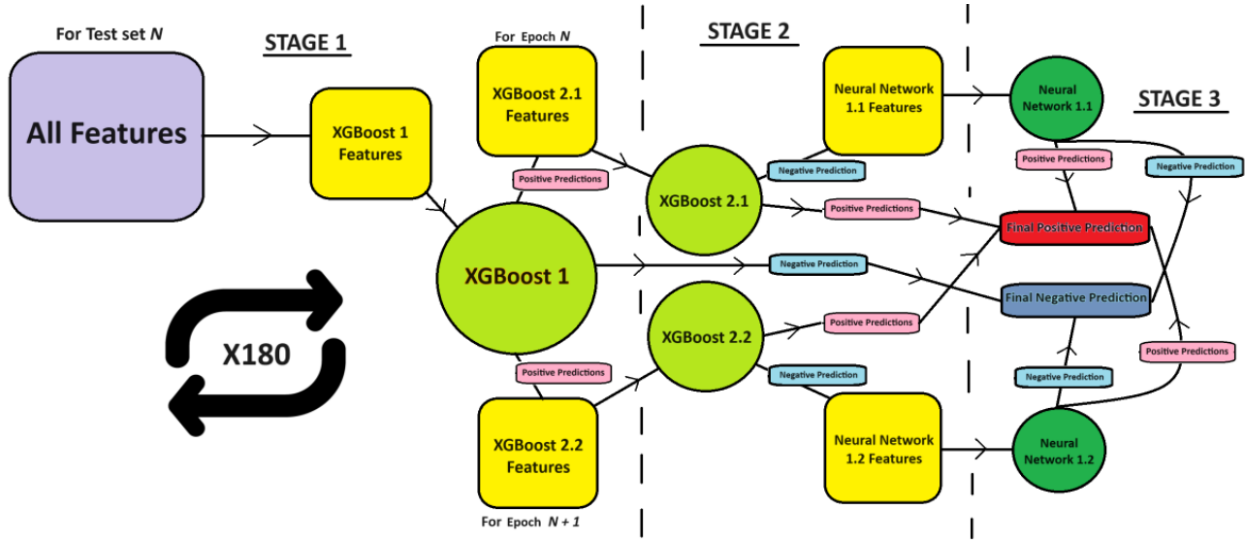


Figure 23: General Workflow of the BPE. Yellow squares represent the particular suite of input features for a specific sub model. The lime green circles represent gradient boost models (using the XGBoost library). The Forest green circles represent neural networks (using TensorFlow library). Pink rectangles are positive predictions, light blue rectangles represent negative predictions. The red rectangle represents the final negative predictions, and dark blue rectangle represents the final positive predictions.

After experimenting with 10 multi-model architectures, the 11th version—the Bi-Path Ensemble (BPE)—proved to be the most effective based on evaluation metrics. The BPE is a multimodal ensemble that combines five fitted gradient boosting models with two fitted neural networks, operating across 180 iterations. The final ensemble prediction for each instance is derived by averaging the output probabilities from all iterations. A set of 180 iterations that are combined to create an ensemble prediction is referred to as a "realization."

Core Structure and Process

The BPE follows a structured three-stage approach:

1. **Stage 1 (XGBoost 1):** Eliminates instances classified as "obviously negative" based on training data.

2. **Stage 2 (XGBoost 2.1 and 2.2):** Eliminates instances classified as "obviously positive."

3. **Stage 3 (Neural Network 1.1 and 1.2):** Classifies the remaining instances by modeling the most subtle relationships between features and labels.

Each stage refines the dataset, ensuring that Stage 3 focuses on the most ambiguous cases. Instances eliminated in earlier stages receive immediate final labels (0 for negative in Stage 1, 1 for positive in Stage 2) and do not proceed further. The terms "obviously positive" and "obviously negative" refer to instances where feature patterns strongly indicate a positive or negative classification, respectively. Exploratory models revealed that gradient boosting and neural networks could maximize either true positives or true negatives but struggled to achieve a balance.

- **Obviously Negative Instances:** These cases exhibited feature values that were highly aligned with negative labels in training data, making them consistently rejected by even the simplest models. They often included areas where environmental conditions strongly suggested the absence of archaeological features (e.g., steep slopes, poor vegetation structure, or known non-cultural formations).
- **Obviously Positive Instances:** These cases were strongly correlated with positive labels, meaning even a weak classifier would identify them with high confidence. These typically included features that exhibited clear structural consistency that aligned with canopy structures of training labels.

Because exploratory models tended to bias toward maximizing either precision or recall, the BPE strategically separates these cases early in the pipeline. This allows the final stage (Stage 3) to focus only on the most ambiguous cases, where subtle patterns in canopy structure or environmental signals may be the deciding factor.

The model selects different features at each stage of the model as a way to diversify the data by varying the suites of features.

- **Stage 1:** Uses all 12 features.
- **Stage 2:** Uses the top 7 features selected based on gain scores from XGBoost 1.
- **Stage 3:** Uses the margin and output probabilities from Stages 1 and 2 as input features.

Pathways for Model Diversity

To enhance prediction robustness, the BPE employs two distinct pathways:

- **Even-numbered iterations:** Prioritize capturing true positives (XGBoost 2.1 and Neural Network 1.1).
- **Odd-numbered iterations:** Prioritize capturing true negatives (XGBoost 2.2 and Neural Network 1.2).

Each instance receives 180 unique output probability scores across iterations, which are averaged to generate the final prediction. Each iteration alters the training set, yet keeps the same testing set.

Rationale behind the architecture:

1. **Diversity:** By altering training sets, validation sets, and feature selection across iteration, the BPE maximizes the range of learned relationships.
2. **Elimination:** By sequentially removing "easy-to-classify" instances, the model dedicates later stages to fine-tuning complex classifications.

Performance and Limitations

The BPE outperformed alternative architectures in both statistical and visual evaluations. However, the model has notable challenges:

- **Computational Demand:** A full ensemble prediction takes ~2 hours, making iterative improvements time-intensive.
- **Hyperparameter Optimization:** Requires tuning for multiple sub-models, necessitating a more automated approach in future versions.

Considering the task is to make predictions on physical space, here is a demonstration of how predictions are assigned at each stage in a spatial demonstration:

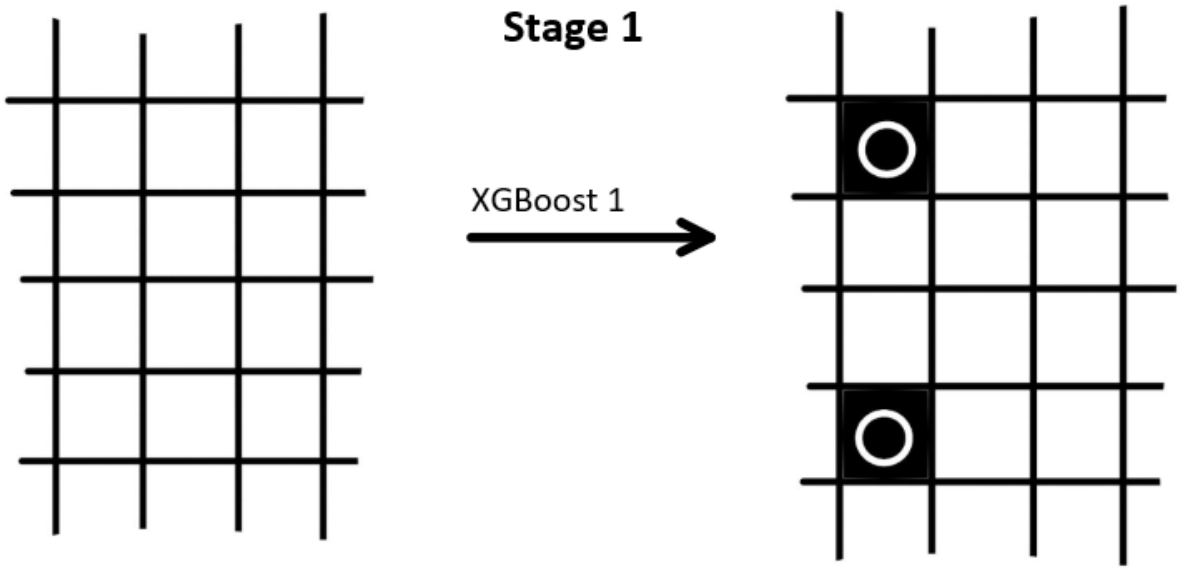


Figure 24: Stage 1 the BPE. The black grids represent the negative instances that were eliminated in this stage.

At stage 1 (figure 24), XGBoost 1 trains on all instances in the prediction dataset (test dataset or labeled dataset) and generates predictions for all instances in that dataset. Instances that are predicted negative are 'eliminated'—and immediately assigned a final predicted label of 0 (negative). Eliminated instances do move on to the next stage.

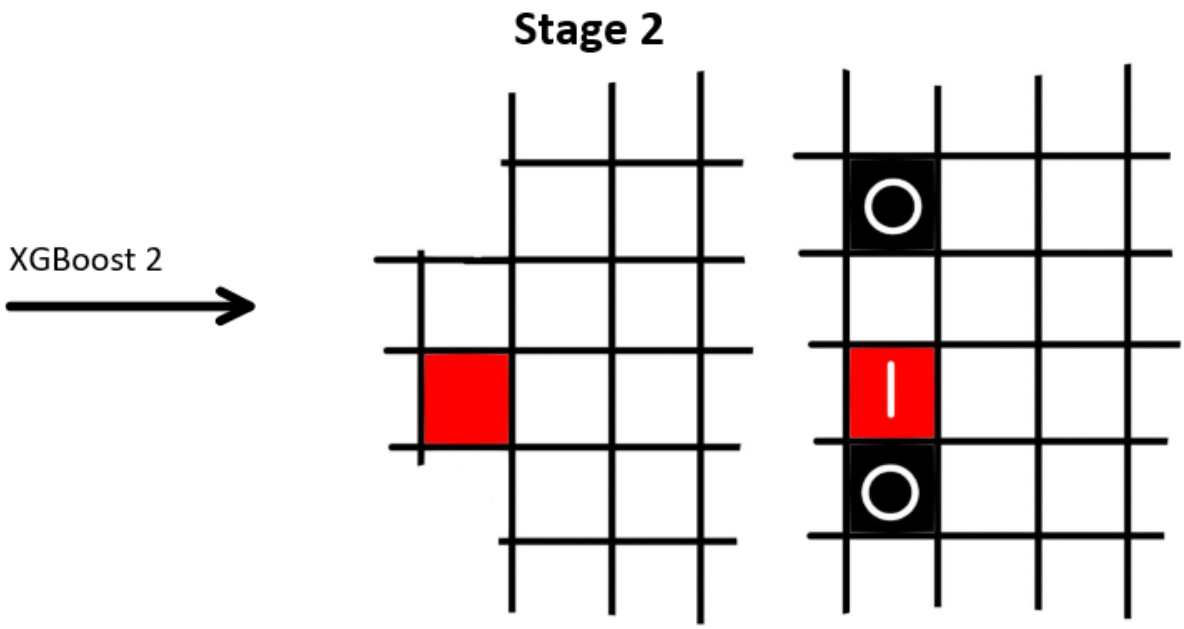


Figure 25: Stage 1 the BPE. The black grids represent the negative instances that were eliminated in this stage.

The remaining instances move on to stage 2 (figure 25) where they are run through XGBoost 2.1 or 2.2, depending on the iteration. Instances predicted to be positive are 'eliminated' –and immediately assigned the final predicted label 1 (positive). Eliminated instances do not move on to stage 3.

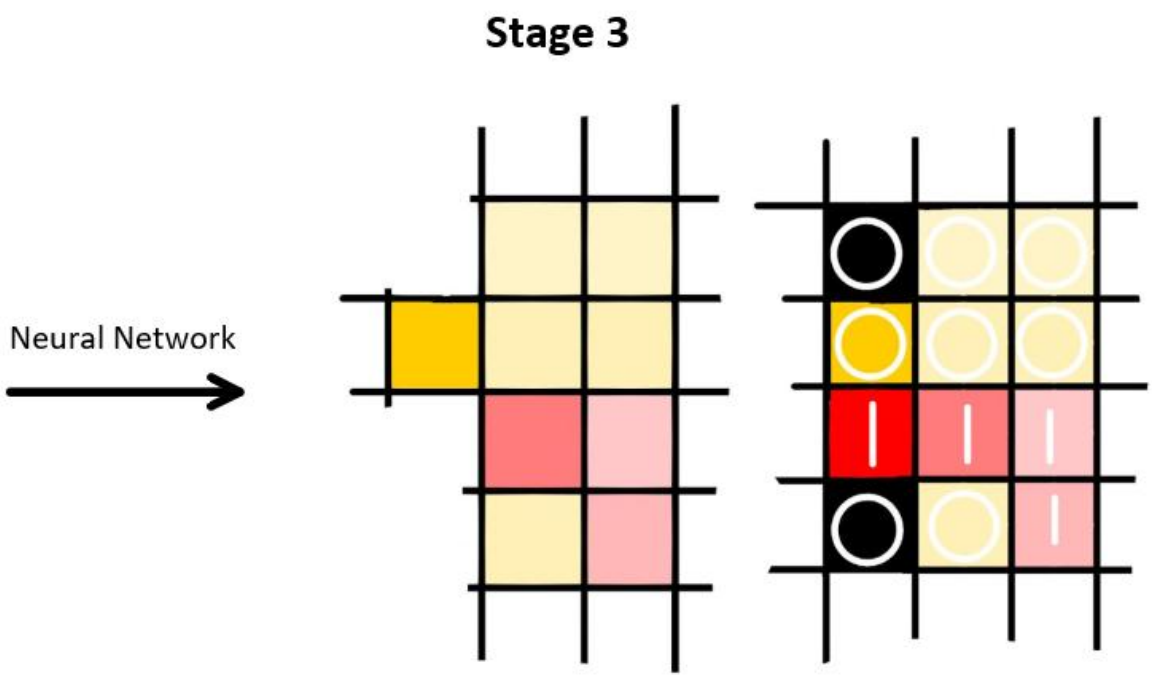


Figure 26: Spatial representation of stage 3 of The BPE. Lighter colors represent instances with output predictions closer to the classifier threshold.

The remaining instances run through stage 3 (figure 26); NN 1.1 or NN 1.2, depending on the iteration. The remaining instances are then given positive or negative labels depending on whether their output probabilistic classifier is greater than or less than the classifier threshold.

BPE's Logic for Angamuco

Theoretically, the BPE model is designed to meet the specific constraints identified for this study by drawing on an ensemble approach that is particularly suited to the complex, fragmented, and imbalanced nature of archaeological data. One of the primary challenges is the site's environmental and cultural heterogeneity—its diverse topography and archaeological typographies—which may make consistent effects between specific archaeological phenomena and canopy structures difficult to find. The BPE's use of multiple weak learners helps to account

for these varied relationships by capturing different perspectives on the data, allowing it to model complex spatial patterns that other single-algorithm models might overlook.

Moreover, the BPE's architecture is particularly effective for dealing with archaeological slow data. This type of data is inherently inconsistent and may suffer from gaps or missing information due to the nature of archaeological research and site formation processes. By using an ensemble method, the model can theoretically adapt to these inconsistencies, ensuring more robust predictions even when data completeness is lacking. This makes it particularly suited to the archaeological context, where uncertainty and incomplete datasets are the norm rather than the exception.

Chapter 3 Summary and Conclusions

The methodological approach presented in this chapter provides a structured process for transforming raw spatial and environmental data into predictive insights about Angamuco's archaeological landscape. By integrating statistical validation techniques with machine learning, the study ensures that label selection is empirically justified and that the predictive models are optimized for performance. The emphasis on multimodal modeling acknowledges the inherent variability in archaeological data and typology and allows the model to more readily adapt to multiplex relationships.

Chapter 4: Results and Discussion

The results of this study indicate that the predictive power of the BPE model varied significantly across different labels. Predictions for Private Non-Terraced (privNT) and Dense Public (densePub) spaces showed promising results, suggesting that the model is capable of identifying and differentiating these features with a reasonable degree of accuracy. However, predictions for Rural Terraces (ruralTerr) yielded notably lower performance, with weak predictive power across several evaluation metrics. This disparity highlights the challenges inherent in predicting more subtle or complex archaeological features, which may be less represented or harder to distinguish from environmental patterns.

This chapter will begin by reviewing the statistical results for each label, providing a quantitative assessment of model performance. Next, I will present the visual outputs, illustrating how the predictions align with the known archaeological features on the ground. Finally, I will discuss the implications of these results, identify enduring drawbacks of the current model, and suggest areas for future research to address these limitations and improve predictive capabilities. This analysis will help contextualize the model's strengths and weaknesses in relation to the archaeological landscape, offering insight into how these findings can inform future efforts in predictive modeling for archaeological site detection.

Rural Terraces

The performance of the ruralTerr label in the BPE model was generally low, with AUC values ranging between 0.56 and 0.60. These modest scores indicate that the model struggled to differentiate between positive and negative instances for this category, suggesting weak predictive power. Given these results, I opted not to conduct a deeper statistical evaluation and

visual evaluations of the ruralTerr label, as the lack of a clear pattern or strong classification performance did not warrant further investigation. While understanding the shortcomings of the model to reliably predict this label compared to the others poses an interesting question, the interpretation of the mechanics of these differences is outside the scope of this research. Instead, emphasis was placed on labels that demonstrated higher predictive reliability.

Dense Public Space

Test Seed	Balanced Accuracy	Kappa	AUC	F1	MCC
42	0.738667	0.325852	0.782502	0.479233	0.370434
43	0.728627	0.285369	0.766231	0.455331	0.345402
44	0.734604	0.333177	0.793886	0.481605	0.369922
45	0.65569	0.298123	0.746212	0.421053	0.298732
45	0.65569	0.298123	0.746395	0.421053	0.298732
46	0.650589	0.327669	0.760488	0.426966	0.330554
47	0.662808	0.280292	0.743748	0.420168	0.286598
48	0.693121	0.319153	0.759062	0.456	0.33055
49	0.691125	0.301744	0.720593	0.44697	0.318002
50	0.709413	0.273099	0.764276	0.442424	0.320062
Average	0.692033	0.30426	0.758339	0.44508	0.326899

Figure 27: Performance of BPE on 8 different realizations. The best results are highlighted in purple.

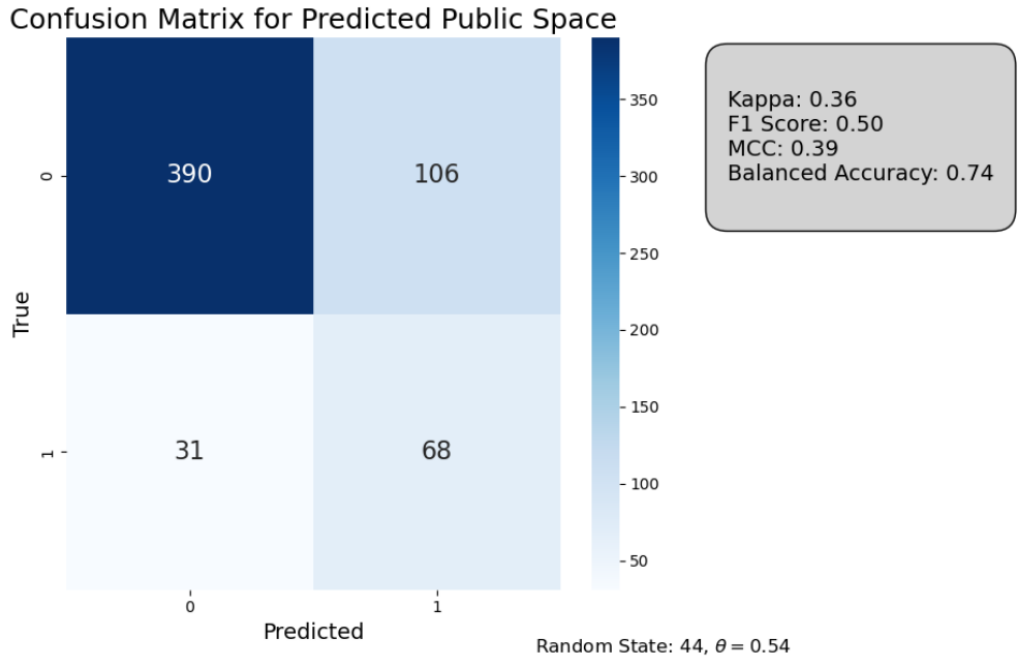


Figure 28: Confusion matrix and evaluation stats for the best prediction (test seed 44).

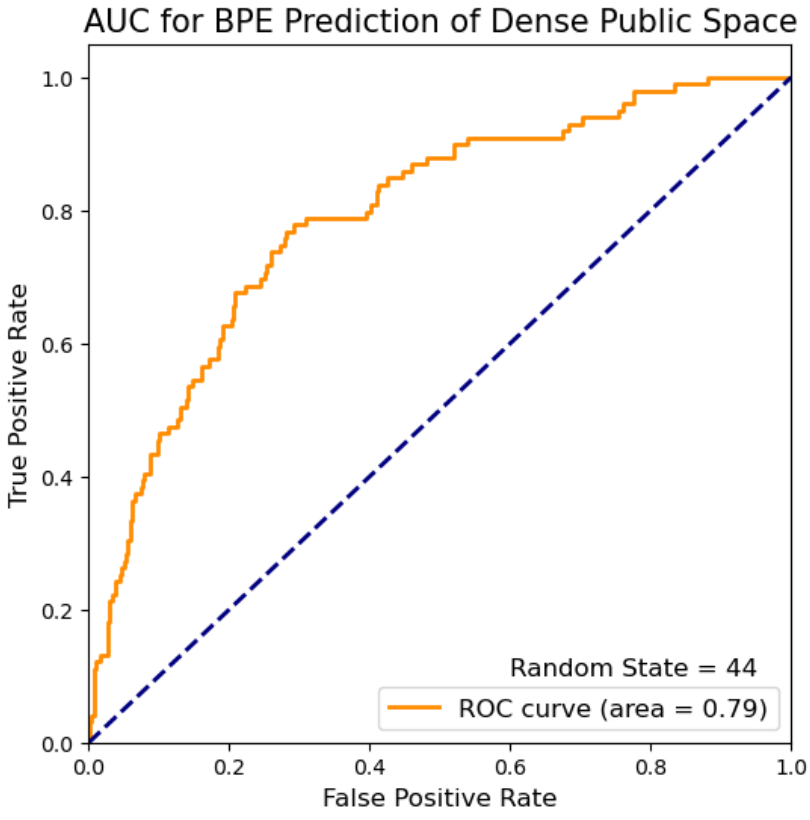


Figure 29: AUC curve for the best prediction (test seed 44).

The BPE model demonstrated strong performance across all evaluation metrics, showcasing its capability to distinguish between dense public space (densePub) areas and non-dense public areas. The balanced accuracy scores ranged from 0.650 to 0.738, indicating that the model was able to correctly predict both positive and negative instances effectively, without bias toward one class.

The Kappa scores ranged from 0.280 to 0.333, suggesting a moderate level of agreement between predicted and actual labels. This indicates that the model has a reasonable level of accuracy, but there is still some room for improvement. The AUC scores, ranging from 0.743 to 0.794, highlight the model's strong ability to separate positive and negative cases. The F1 scores, which balance precision and recall, ranged from 0.421 to 0.481. These values suggest that the model was effective at both identifying true positive cases (high recall) and minimizing false positives (high precision), though there is still potential to enhance its precision-recall trade-off. Finally, the MCC scores ranged from 0.285 to 0.370, pointing to a moderate positive correlation between predicted and true labels. This indicates that while there is a reasonable relationship between the model’s predictions and actual outcomes, further model refinement could yield stronger correlations.

These results collectively suggest that the BPE model performs well in predicting densePub locations. However, there is a lot of room for further optimization to improve the balance between precision, recall, and discriminatory power.

Defining Almost Correct

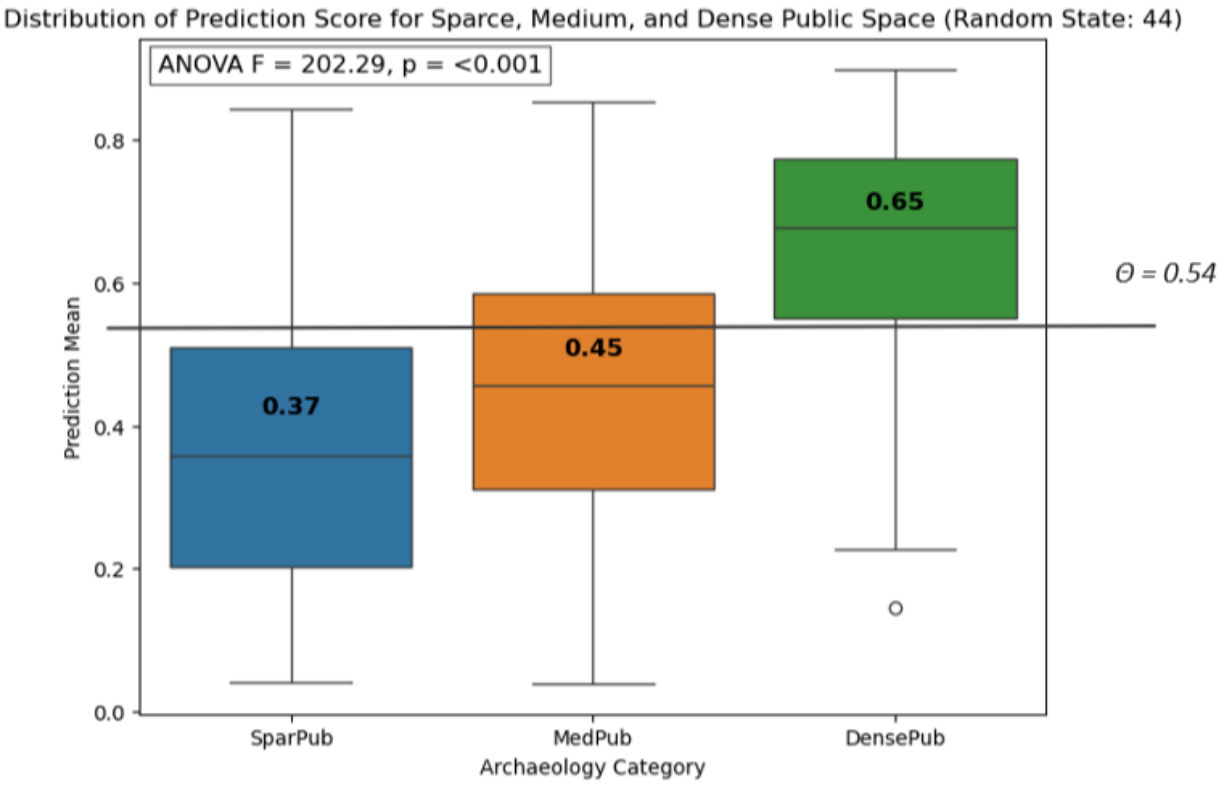


Figure 30: Mean (ensemble) prediction scores assigned to dense public space label category

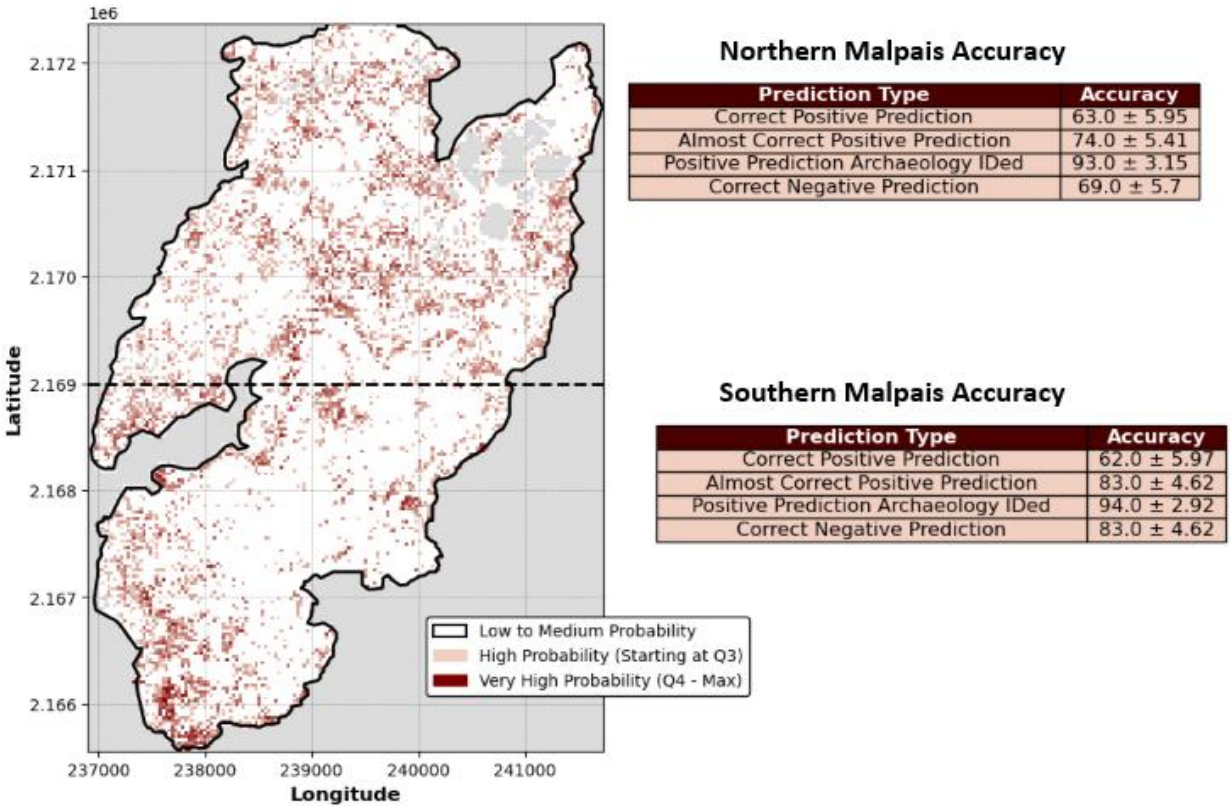
While an ANOVA test revealed significant differences in the mean prediction scores between the densePub, medPub, and sparsePub labels, further investigation into the data distribution showed that many of the false positives were attributed to the medPub and sparsePub labels (Figure 30). Given this, I decided to categorize these instances as "almost correct." This classification was based on the understanding that medPub and sparsePub predictions, while not perfect, were still relatively close to the densePub predictions in terms of spatial characteristics. By including these labels in the "almost correct" category, I aimed to capture the nuances of model behavior where the predicted values were not entirely accurate but were still closely related to the true positive predictions. This decision enhances the robustness of the evaluation by accounting for overlapping label characteristics and the inherent complexity of the model's predictions.

South Site Visual Evaluation:

The densePub model's performance on the south site shows strong results for positive predictions. Correct positive predictions had an accuracy of 62.0 ± 5.97, while almost correct positive predictions were notably higher at 83.0 ± 4.62. The model successfully identified archaeological features in 94.0 ± 2.92 of positive predictions, suggesting it can detect relevant archaeological phenomena even when not perfectly classified as densePub. Correct negative predictions had an accuracy of 83.0 ± 4.62, indicating reliable differentiation between densePub and non-densePub instances.

North Site Visual Evaluation:

For the north site, accuracy for correct positive predictions was similar to the south at 63.0 ± 5.95, with a slightly lower rate for almost correct positive predictions (74.0 ± 5.41). Like the south, positive predictions identified archaeological features at 93.0 ± 3.15, demonstrating the model's consistency in detecting archaeological phenomena. Correct negative predictions were slightly lower than in the south, at 69.0 ± 5.7, showing more variability in distinguishing non-densePub features. Overall, the model performs similarly on both sites, with a slight decrease in accuracy for the north region.



Northern Malpais Accuracy

Prediction Type	Accuracy
Correct Positive Prediction	63.0 ± 5.95
Almost Correct Positive Prediction	74.0 ± 5.41
Positive Prediction Archaeology IDed	93.0 ± 3.15
Correct Negative Prediction	69.0 ± 5.7

Southern Malpais Accuracy

Prediction Type	Accuracy
Correct Positive Prediction	62.0 ± 5.97
Almost Correct Positive Prediction	83.0 ± 4.62
Positive Prediction Archaeology IDed	94.0 ± 2.92
Correct Negative Prediction	83.0 ± 4.62

Figure 31: BPE predicted dense public space, visual evaluation results

Private Space (No Terrace)

Test Seed	Balanced Accuracy	Kappa	AUC	F1	MCC
42	0.612661	0.291533	0.740482	0.321429	0.309481
45	0.620798	0.280532	0.722803	0.31746	0.285408
46	0.639761	0.279521	0.728083	0.324324	0.279521
49	0.600044	0.267621	0.692386	0.296296	0.290034

50	0.660443	0.351972	0.761939	0.38806	0.354147
51	0.731958	0.239912	0.789257	0.310811	0.287589
53	0.704785	0.272195	0.785479	0.333333	0.294671
54	0.573913	0.208264	0.711421	0.235294	0.235513
Average	0.643045	0.273944	0.741481	0.315876	0.292046

Figure 32: Performance of BPE on 8 different realizations. The best results are highlighted in purple.

The BPE model demonstrates moderate predictive performance across key evaluation metrics. Balanced accuracy scores range from 57% to 73%, indicating that the model is capable of distinguishing between privNT and non-privNT features with reasonable effectiveness. The AUC values, which fall between 0.64 and 0.79, suggest that the model has a fair ability to rank true positive predictions higher than false positives, though there is room for improvement in overall discrimination.

Kappa scores, ranging from 0.2 to 0.35, reflect moderate agreement between the predicted and actual labels, suggesting that the model captures meaningful patterns but does not achieve strong classification confidence. The F1 and MCC scores reinforce this, showing that while the model maintains a balance between precision and recall, its overall predictive power remains moderate when predicting the locations of unterraced private space.

Defining Almost Correct

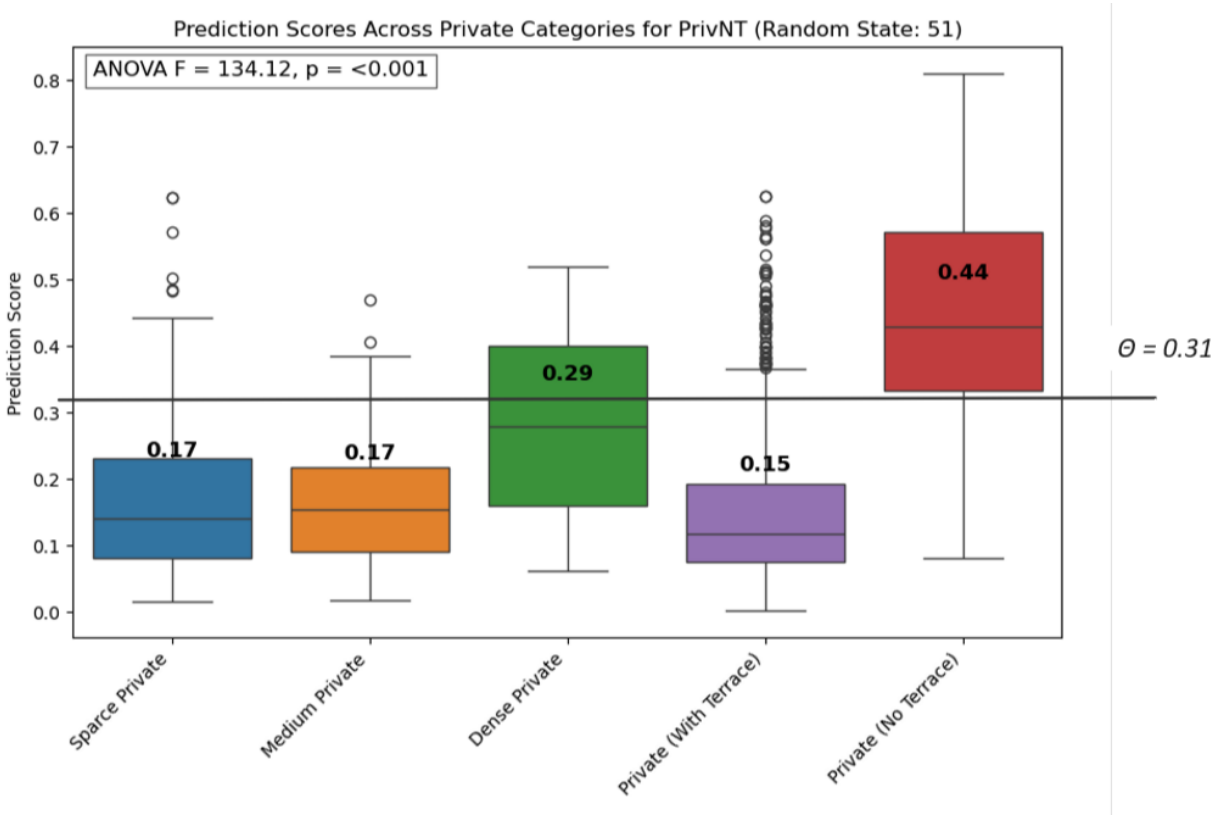


Figure 33: Mean (ensemble) prediction scores assigned to private space label categories

As with densePub, an ANOVA test revealed significant differences in the mean prediction scores between the privNT and other labels, further investigation into the data distribution showed that many of the false positives were attributed to the densePriv label. Given this, I decided to categorize these instances as "almost correct."

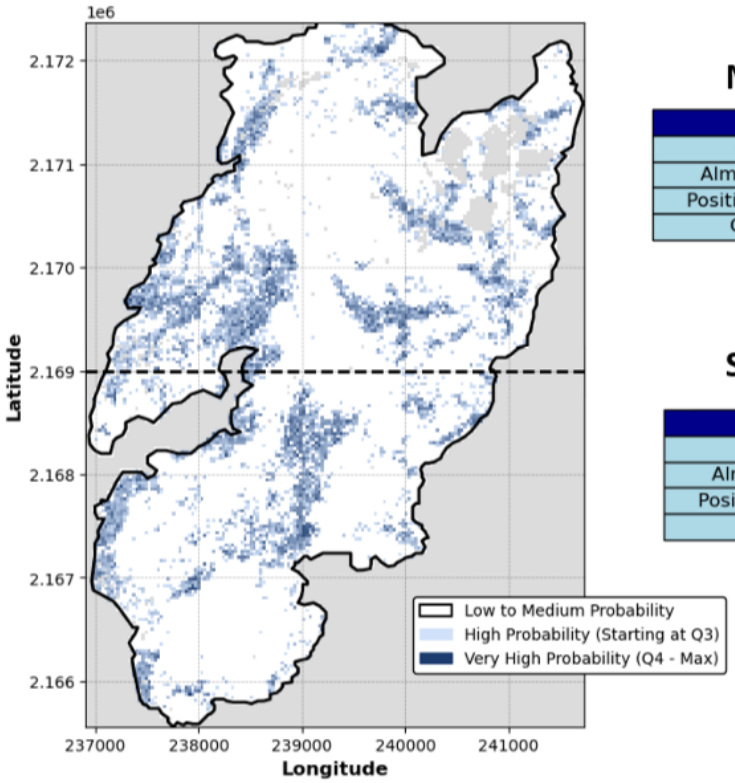
South Site Visual Evaluation

On the south site, the model showed reasonable performance for correct positive predictions at 53.0 ± 6.13, indicating that a little more than half of the positive predictions were accurate. However, the rate of almost correct positive predictions was higher at 73.0 ± 5.46, reflecting that many positive predictions were related to the privNT category but not perfectly identified. The model was highly successful in identifying archaeological features in 93.0 ± 3.14

of positive predictions, suggesting that it can detect relevant archaeological phenomena even if not perfectly categorized as privNT. Correct negative predictions were at 81.0 ± 4.82, demonstrating a relatively high accuracy in differentiating privNT from non-privNT instances.

North Site Visual Evaluation:

For the north site, the accuracy for correct positive predictions was lower at 39.0 ± 6.02, indicating a less accurate detection of privNT features compared to the south site. Almost correct positive predictions were also lower at 66.0 ± 5.84, showing that the model struggled more in correctly identifying privNT features in this region. However, the model still managed to identify archaeological features in 82.0 ± 4.74 of positive predictions, which shows some degree of archaeological relevance even in misclassified instances. Correct negative predictions were relatively strong at 83.0 ± 4.63, showing a solid ability to identify non-privNT instances accurately. Overall, the model's performance was lower in the north site, particularly for correctly identifying privNT features.



Northern Malpais Accuracy

Prediction Type	Accuracy
Correct Positive Prediction	39.0 ± 6.02
Almost Correct Positive Prediction	66.0 ± 5.84
Positive Prediction Archaeology IDed	82.0 ± 4.74
Correct Negative Prediction	83.0 ± 4.63

Southern Malpais Accuracy

Prediction Type	Accuracy
Correct Positive Prediction	53.0 ± 6.13
Almost Correct Positive Prediction	73.0 ± 5.46
Positive Prediction Archaeology IDed	93.0 ± 3.14
Correct Negative Prediction	81.0 ± 4.82

Figure 34: BPE predicted privNT, visual evaluation results

Discussion: Statistical and Visual Evaluation of *densepub* and *privNT* Labels

The analysis of the *densepub* and *privNT* labels presents a nuanced comparison of spatial and environmental features related to archaeological landscape prediction. The relationship between these labels and the model results offers important insights into how tree canopy structure can aid in predicting the location of archaeological features, such as terraces and ancient residential areas.

The results of this study suggest that the BPE model holds significant promise in leveraging canopy structure to predict the locations of specific archaeological typologies, though its highest accuracy use case remains identifying archaeology more generally. The strong

performance of *densepub*—which had the highest PerMANOVA F-statistic and the best BPE evaluation metrics—demonstrates that certain archaeological typologies exert a clear and consistent influence on canopy structure, allowing for highly accurate predictions. Likewise, the *privNT* label, despite lacking statistical significance in perMANOVA, still exhibited reasonably strong predictive performance in the BPE model, suggesting that even subtle canopy variations associated with certain archaeological patterns can be detected.

However, the case of *ruralTerr* illustrates the complexity of this relationship. While perMANOVA identified it as statistically significant, the BPE model struggled to generate strong predictive results, likely due to the dispersed and variable nature of rural terraces. This suggests that while perMANOVA can highlight broad statistical relationships, it does not necessarily translate into strong BPE performance.

The visual evaluation results provide further insight into the nuances of BPE's predictive capabilities. Many false positives in the test datasets corresponded to related labels (*medPub*, *densePriv*, etc.), meaning that while the BPE model did not always predict the exact label, it often identified a closely related archaeological phenomenon. This pattern indicates that the model is honing in on meaningful archaeological distinctions but at a slightly broader level than the assigned labels. In other words, while the BPE does not always precisely classify individual typologies, it does appear to differentiate between broad archaeological patterns in a way that is more refined than a simple archaeology/no-archaeology distinction.

Taken together, these findings suggest that the BPE model has significant potential for improving archaeological survey efforts by identifying specific types of archaeological phenomena, albeit with some level of generalization. Its highest-confidence predictions align well with known archaeological typologies, but its predictive boundaries remain somewhat fluid,

with overlap among closely related classes. As such, the BPE is best suited for identifying archaeology in general, while still providing valuable insights into the likely nature of that archaeology. Future work could refine these predictions by integrating additional contextual variables—such as topography, soil composition, or historical land use data—to further enhance specificity and accuracy.

Broadly, this work emphasizes the value of integrating environmental proxies into emerging AASMs to enhance landscape-level predictions. Furthermore, I hope the successes of this study demonstrates that creative applications of machine learning algorithms can tackle the complexities of archaeological data.

Drawbacks and Future Considerations

While the BPE model has demonstrated promising results in predicting archaeological features based on tree canopy structure, several drawbacks must be addressed to enhance its overall effectiveness and reliability. One major limitation is the computational efficiency of the model. As it stands, the model requires considerable processing power and time to analyze large datasets, which could present challenges when scaling the analysis to larger sites or applying it in real-world archaeological surveys. Optimization of the model's algorithms is necessary to improve processing speed without sacrificing accuracy. This will be crucial for practical applications where time and resources are often constrained.

Moreover, while the BPE has shown potential in identifying general archaeological patterns, its predictions must be validated through ground-truthing, as noted by Reese-Taylor (2016). In particular, the possibility of anomalies in the DEM datasets—such as cultural features masked by vegetation or natural topographic features that resemble archaeological structures—

could introduce inaccuracies into the predictions. As Reese-Taylor suggests, ground-truthing efforts are essential to verify the model's effectiveness and ensure that the detected anomalies correspond to real archaeological phenomena. Without this validation, the model's results remain speculative, underscoring the need for further empirical testing to establish the robustness of the model's predictions.

Another challenge lies in the accuracy assessments of the northern area of the site, which is home to a diverse range of typologies spanning different phases of habitation. According to Harris (2019), the archaeological features in this area may be subject to greater interpretive variability, as different cultural and historical contexts overlap (Harris, 2019). The presence of multiple layers of occupation in this region means that the visual evaluation of BPE predictions is somewhat subjective, as it requires disentangling the effects of later construction or agricultural activities from earlier features. This complexity makes it difficult to fully assess the model’s accuracy in these regions without further refinement of the model’s ability to differentiate between overlapping temporal and spatial features.

Given these issues, it is important to recognize that the effectiveness of the BPE model, at this current stage, remains largely theoretical. While the model's results are promising and provide valuable insights into the potential for using tree canopy structure to predict archaeological features, the lack of ground-truthing, computational inefficiencies, and interpretive challenges in areas with overlapping typologies suggest that the model’s full applicability and reliability have yet to be established. Therefore, while the BPE holds considerable promise for enhancing archaeological survey methods, its current implementation requires refinement and further validation before it can be considered a fully reliable tool for archaeological prediction.

Future Research Directions

While this study demonstrates the potential of BPE for predicting archaeological features in Angamuco, further research is necessary to fully understand the ecological mechanisms driving these predictions. One key area of investigation is how archaeological materials influence tree canopy structure at the site. While prior research has shown that terraces and built environments can affect vegetation growth, the specific interactions between soil composition, moisture retention, and tree species in Angamuco remain unclear. A more detailed ecological study integrating soil chemistry, root structure analysis, and long-term vegetation surveys would help clarify these relationships and improve future predictive modeling efforts.

Additionally, the northern site presents a different archaeological signature compared to other areas analyzed in this study. It contains distinct material compositions and potentially different land-use histories, raising questions about whether the accuracy assessments applied here remain valid across all regions of Angamuco. Further validation efforts, including ground-truthing and refined classification models, would help assess whether the BPE approach maintains predictive consistency across varying archaeological contexts.

Beyond Angamuco, testing BPE on other archaeological malpaís landscapes in the Pátzcuaro Basin and beyond would provide crucial insight into its broader applicability. Malpaís sites present unique preservation and detection challenges, and extending this methodology to additional locations could refine the model's robustness. Expanding the study area would also provide a comparative dataset to assess whether the observed canopy-archaeology relationships are site-specific or part of a broader regional pattern.

Chapter 5: Conclusion

This study aimed to explore the potential of tree canopy structure as a predictive tool for identifying archaeological features, particularly in areas where traditional survey methods may be impractical. The central hypothesis—that canopy patterns can reflect past human activity in predictable ways—was supported by the results, which demonstrated the BPE model's ability to predict areas with significant archaeological typologies, such as dense public spaces, with high accuracy.

The findings highlight that the BPE model is particularly effective at identifying large-scale, highly structured areas like dense public spaces, where canopy patterns are most strongly influenced by infrastructure. However, the model struggled to predict the locations of other land-use typologies, such as rural terraces. Despite this, the BPE consistently differentiated between general archaeological and non-archaeological areas, reinforcing the utility of tree canopy data in identifying regions of interest for further archaeological investigation.

These results underscore the value of using remote sensing and machine learning to conduct large-scale archaeological surveys. This approach offers a scalable and efficient way to identify areas where human-modified landscapes are likely to be found, reducing the reliance on labor-intensive ground surveys. The model's ability to predict general archaeological patterns also suggests that it can serve as a valuable first step in prioritizing areas for high-resolution feature extraction and more targeted analysis.

While the study demonstrates promising results, some limitations should be considered. The variability in predicting certain features, particularly rural terraces, suggests that further refinement is needed to improve the accuracy of the model across a wider range of

archaeological typologies. Factors such as the inconsistent nature of some features and the potential influence of additional environmental variables (e.g., soil composition, topography) may be contributing to these challenges.

Future work should aim to enhance the model's specificity by incorporating contextual data, such as historical land use or topographic information, to refine predictions. Additionally, further research could explore the integration of higher-resolution datasets to improve the model's capacity for distinguishing between closely related archaeological patterns. As remote sensing and machine learning techniques continue to evolve, there is substantial potential for improving the accuracy and applicability of canopy-based archaeological survey methods.

In conclusion, this research provides valuable insights into the use of tree canopy data for archaeological detection, offering a promising tool for large-scale surveys and conservation efforts. By refining the model and incorporating additional contextual variables, future studies could significantly improve the precision of archaeological predictions, expanding the potential of automated archaeological survey methods in complex environments.

Works Cited

Albrecht, C. M., Fisher, C., Freitag, M., Hamann, H. F., Pankanti, S., Pezzutti, F., & Rossi, F. (2019). Learning and Recognizing Archeological Features from LiDAR Data. *2019 IEEE International Conference on Big Data (Big Data)*, 5630–5636. https://doi.org/10.1109/BigData47090.2019.9005548

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*(1), 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

Ben-Romdhane, H., Francis, D., Cherif, C., Pavlopoulos, K., Ghedira, H., & Griffiths, S. (2023). Detecting and Predicting Archaeological Sites Using Remote Sensing and Machine Learning—Application to the Saruq Al-Hadid Site, Dubai, UAE. *Geosciences*, *13*(6), Article 6. https://doi.org/10.3390/geosciences13060179

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Bickler, S. H. (2021). Machine Learning Arrives in Archaeology. *Advances in Archaeological Practice*, *9*(2), 186–191. https://doi.org/10.1017/aap.2021.6

Chase, A. F., Chase, D. Z., Fisher, C. T., Leisz, S. J., & Weishampel, J. F. (2012). Geospatial revolution and remote sensing LiDAR in Mesoamerican archaeology. *Proceedings of the National Academy of Sciences*, *109*(32), 12916–12921. https://doi.org/10.1073/pnas.1205198109

Cohen, A. S. (2016). *Creating an Empire: Local Political Change at Angamuco, Michoacan, Mexico* [PhD Thesis].

Dalponte, M., & Coomes, D. A. (2016). Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods in Ecology and Evolution*, *7*(10), 1236–1245. https://doi.org/10.1111/2041-210X.12575

Davis, D. (2021). Theoretical repositioning of automated remote sensing archaeology: Shifting from features to ephemeral landscapes. *Journal of Computer Applications in Archaeology*, *4*(1), 94–109.

Davis, D. S., Sanger, M. C., & Lipo, C. P. (2019). Automated mound detection using lidar and object-based image analysis in Beaufort County, South Carolina. *Southeastern Archaeology*, *38*(1), 23–37. https://doi.org/10.1080/0734578X.2018.1482186

Dey, V., Zhang, Y., Zhong, M., & Salehi, B. (2011). Image segmentation techniques for urban land cover segmentation of VHR imagery: Recent developments and future prospects. *VIVEK DEY*, 34.

Dietterich, T. G. (2002). Ensemble learning. *The Handbook of Brain Theory and Neural Networks*, *2*(1), 110–125.

Dorison, A. (2022). Ancient Agriculture on Lava Flows: Using LiDAR and Soil Science to Reassess Pre-Hispanic Farming on Malpaís Landforms in West Mexico. *Journal of Ethnobiology*, *42*(2), 131–151. https://doi.org/10.2993/0278-0771-42.2.131

Ebert, J. I. (1984). 5—Remote Sensing Applications in Archaeology. In M. B. Schiffer (Ed.), *Advances in Archaeological Method and Theory* (pp. 293–362). Academic Press. https://doi.org/10.1016/B978-0-12-003107-8.50010-4

El Naqa, I., & Murphy, M. J. (2015). What Is Machine Learning? In I. El Naqa, R. Li, & M. J. Murphy (Eds.), *Machine Learning in Radiation Oncology: Theory and Applications* (pp. 3–11). Springer International Publishing. https://doi.org/10.1007/978-3-319-18305-3_1

Estrada-Belli, F., & Koch, M. (2007). Remote Sensing and GIS Analysis of a Maya City and Its Landscape: Holmul, Guatemala. In J. Wiseman & F. El-Baz (Eds.), *Remote Sensing in Archaeology* (pp. 263–281). Springer. https://doi.org/10.1007/0-387-44455-6_11

Fisher, C. T., & Leisz, S. J. (2013). New Perspectives on Purépecha Urbanism Through the Use of LiDAR at the Site of Angamuco, Mexico. In D. C. Comer & M. J. Harrower (Eds.), *Mapping Archaeological Landscapes from Space* (pp. 199–210). Springer. https://doi.org/10.1007/978-1-4614-6074-9_16

Freeland, T., Heung, B., Burley, D. V., Clark, G., & Knudby, A. (2016). Automated feature extraction for prospection and analysis of monumental earthworks from aerial LiDAR in the Kingdom of Tonga. *Journal of Archaeological Science*, *69*, 64–74. https://doi.org/10.1016/j.jas.2016.04.011

Guyot, A., Lennon, M., Lorho, T., & Hubert-Moy, L. (2021). Combined Detection and Segmentation of Archeological Structures from LiDAR Data Using a Deep Learning Approach. *Journal of Computer Applications in Archaeology*, *4*(1), 1. https://doi.org/10.5334/jcaa.64

Harris, E. (2019). *Viewed from above: Extracting the built environment from the ancient Purépecha site of Angamuco through development of a new methodology*. https://hdl.handle.net/10217/199740

Hesse, R. (2009). *Extraction of archaeological features from high-resolution LIDAR data*. https://www.academia.edu/1045403/Extraction_of_archaeological_features_from_high_resolution_LIDAR_data

Hightower, J. N., Butterfield, A. C., & Weishampel, J. F. (2014). Quantifying Ancient Maya Land Use Legacy Effects on Contemporary Rainforest Canopy Structure. *Remote Sensing*, *6*(11), Article 11. https://doi.org/10.3390/rs61110716

Hosking, J. R. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *52*(1), 105–124.

Howey, M. C. L., Palace, M. W., & McMichael, C. H. (2016). Geospatial modeling approach to monument construction using Michigan from A.D. 1000–1600 as a case study. *Proceedings of the National Academy of Sciences*, *113*(27), 7443–7448. https://doi.org/10.1073/pnas.1603450113

ICOMOS. (2019). *ICOMOS Declares a Climate Emergency—International Council on Monuments and Sites*. https://www.icomos.org/en/focus/climate-change/85740-icomos-declares-a-climate-emergency

INFyS. (2015). *Composicion y Estuctura de Inventario Nacional Forestal y de Suelo* [Dataset]. INFyS.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Lecun, Y. (2001). *A Theoretical Framework for Back-Propagation*.

Leisz, S. J. (2013). An Overview of the Application of Remote Sensing to Archaeology During the Twentieth Century. In D. C. Comer & M. J. Harrower (Eds.), *Mapping Archaeological Landscapes from Space* (pp. 11–19). Springer. https://doi.org/10.1007/978-1-4614-6074-9_2

McGovern, T. H. (2018). Burning Libraries: A Community Response. *Conservation and Management of Archaeological Sites*, *20*(4), 165–174. https://doi.org/10.1080/13505033.2018.1521205

Mlekuž, D. (2013). Skin Deep: LiDAR and Good Practice of Landscape Archaeology. *Natural Science in Archaeology*, 113–129.

Moran, C. J., Rowell, E. M., & Seielstad, C. A. (2018). A data-driven framework to identify and compare forest structure classes using LiDAR. *Remote Sensing of Environment*, *211*, 154–166.

Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., & Weng, Q. (2011). Per-pixel vs. Object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, *115*(5), 1145–1161. https://doi.org/10.1016/j.rse.2010.12.017

Peripato, V., Levis, C., Pitman, N. C. A., Junqueira, A. B., Trindade, T. B., de Almeida, F. O., Tamanaha, E. K., Maezumi, S. Y., Braga, J. R. G., Campanharo, W. A., Cassol, H. L. G., Leal, P. R., de Assis, M. L. R., da Silva, A. M., Phillips, O. L., Costa, F. R. C., Flores, B. M., Magnusson, W. E., Valderrama Sandoval, E. H., … Ramirez Arevalo, F. (2023). More than 10,000 pre-Columbian earthworks are still hidden throughout Amazonia. *Science (American Association for the Advancement of Science)*, *382*(6666), 103–109. https://doi.org/10.1126/science.ade2541

Pollard, H. P. (1980). Central Places and Cities: A Consideration of the Protohistoric Tarascan State. *American Antiquity*, *45*(4), 677–696. https://doi.org/10.2307/280141

Pollard, H. P. (2008). A Model of the Emergence of the Tarascan State. *Ancient Mesoamerica*, *19*(2), 217–230. https://doi.org/10.1017/S0956536108000369

Pollard, H. P., Smith, M. E., & Berdan, F. F. (2003). The tarascan empire. In *The Oxford Handbook of Mesoamerican Archaeology*.

Rafuse, D. J. (2021). A Maxent Predictive Model for Hunter-Gatherer Sites in the Southern Pampas, Argentina. *Open Quaternary*, *7*(1).

Reese-Taylor, K., Hernández, A. A., Esquivel, F. C. A. F., Monteleone, K., Uriarte, A., Carr, C., Acuña, H. G., Fernandez-Diaz, J. C., Peuramaki-Brown, M., & Dunning, N. (2016). Boots on the Ground at Yaxnohcah. *Advances in Archaeological Practice*, *4*(3), 314.

Reyes-Guzmán, N., Siebe, C., Chevrel, M. O., Pereira, G., Mahgoub, A. N., & Böhnel, H. (2023). Holocene volcanic eruptions of the Malpaís de Zacapu and its pre-Hispanic settlement history. *Ancient Mesoamerica*, *34*(3), 712–727. https://doi.org/10.1017/S095653612100050X

Schapire, R. E. (2003). The Boosting Approach to Machine Learning: An Overview. In D. D. Denison, M. H. Hansen, C. C. Holmes, B. Mallick, & B. Yu (Eds.), *Nonlinear Estimation and Classification* (pp. 149–171). Springer. https://doi.org/10.1007/978-0-387-21579-2_9

Silva, C. A., Hudak, A. T., Vierling, L. A., Loudermilk, E. L., O'Brien, J. J., Hiers, J. K., Jack, S. B., Gonzalez-Benecke, C., Lee, H., Falkowski, M. J., & Khosravipour, A. (2016). Imputation of Individual Longleaf Pine (*Pinus palustris* Mill.) Tree Attributes from Field and LiDAR Data. *Canadian Journal of Remote Sensing*, *42*(5), 554–573. https://doi.org/10.1080/07038992.2016.1196582

Solinis-Casparius, R. (2019). *The Role of Road Networks in Social Definition and Integration of Angamuco, Michoacan (250–1530 CE)* [Ph.D., University of Washington]. https://www.proquest.com/pqdtglobal/docview/2307191568/abstract/4D2D7080A0DC4698PQ/1

Urquhart, K. R. (2015). *The Ireta: A model of political and spatial organization of P'urépecha cities* [M.A., Colorado State University]. https://www.proquest.com/docview/1717085003/abstract/A867B61A819749D5PQ/1

Vaart, W. B. V. der, & Lambers, K. (2019). *Learning to Look at LiDAR: The Use of R-CNN in the Automated Detection of Archaeological Objects in LiDAR Data from the Netherlands* (1). *2*(1), Article 1. https://doi.org/10.5334/jcaa.32

Valderrama, L., España, M., Frederic, B., Sánchez-Vargas, N., & Saenz-Romero, C. (2014). Capacity of phenological data derived from CYCLOPES LAI for the year 2000 to distinguish land cover types in the State of Michoacán, Mexico. *Revista Chapingo Serie Ciencias Forestales y Del Ambiente*, *XX*. https://doi.org/10.5154/r.rchscfa.2013.08.025

Vaughn, S., & Crawford, T. (2009). A predictive model of archaeological potential: An example from northwestern Belize. *Applied Geography*, *29*(4), 542–555. https://doi.org/10.1016/j.apgeog.2009.01.001

Wiseman, J., & El-Baz, F. (2007). Introduction. In *Remote Sensing in Archaeology* (pp. 1–8).