Advances in Archaeological Practice (2026), **14**, 179–197
doi:10.1017/aap.2025.10130

SAA
Society for American Archaeology

ARTICLE

Using Simulated Training Data to Locate Archaeological Sites with Machine Learning

Katherine Peck[1,2], Claudine Gravel-Miguel[1,2], Grant Snitker[2,3], and Matthew Helmer[4]

[1]Department of Anthropology, University of New Mexico, Albuquerque, NM, USA; [2]Cultural Resource Sciences Program, New Mexico Consortium, Los Alamos, NM, USA; [3]School of Social Sciences, Utah State University, Logan, UT, USA and [4]Department of Geography and Anthropology, Louisiana State University, Baton Rouge, LA, USA

This research article was awarded an Open Materials badge for transparent practices. See the Data Availability Statement for details.

(Received 6 May 2025; revised 9 September 2025; accepted 9 September 2025; first published online 27 April 2026)

Abstract

Archaeologists have demonstrated the value of deep learning models for detecting archaeological objects in lidar data. As landscape-level projects become the norm, archaeological data derived from deep learning predictions can be integrated into these initiatives through coupled natural-cultural landscapes planning. However, the paucity of archaeological training datasets limits the application of deep learning models to relatively common and well-documented object classes. Using procedurally generated training datasets may be one approach to overcome this bottleneck. To test the efficacy of procedural generation for developing deep learning training data, we trained models to detect a novel object class (hypothesized historic tar kilns) in the Kisatchie National Forest in Louisiana. We developed two procedural generation approaches to embed simulated archaeological objects in a lidar-derived DEM and used these datasets to train deep learning (Mask R-CNN) models. We then evaluated model predictions within lidar-derived visualizations and during field survey. Our trained models detected targets with high recall but low precision. Field investigation suggested that the objects were not tar kilns but a different historic feature class. This study suggests that models trained on simulated objects are a useful addition to lidar analysis tool kits and can be directly integrated into archaeological field investigation workflows.

Resumen

La arqueología ha demostrado el valor de los modelos de aprendizaje profundo ("deep learning" en inglés) para detectar restos arqueológicos en datos lidar. A medida que los proyectos a nivel paisajístico se generalizan en la praxis arqueológica, los datos derivados de las predicciones del aprendizaje profundo pueden integrarse en estas iniciativas mediante la planificación conjunta de paisajes naturales y culturales. Sin embargo, la escasez de conjuntos de datos de entrenamiento específicamente arqueológico limita la aplicación de los modelos de aprendizaje profundo a clases de objetos relativamente comunes y/o bien documentados. El uso de conjuntos de datos de entrenamiento generados procedimentalmente puede ser un enfoque para superar este cuello de botella. Para probar la eficacia de la generación por procedimientos para desarrollar datos de entrenamiento de aprendizaje profundo, hemos entrenado modelos para detectar una nueva clase de objetos (hipotéticos hornos de brea de época histórica) en el Kisatchie National Forest, Louisiana. Desarrollamos dos enfoques de generación procedimental para incrustar objetos arqueológicos simulados en un DEM derivado de lidar y utilizamos estos conjuntos de datos para entrenar modelos de aprendizaje profundo (mask region-based convolutional neural network; Mask R-CNN). A continuación, evaluamos las predicciones del modelo dentro de las visualizaciones derivadas de lidar y durante el estudio de campo. Nuestros modelos entrenados detectaron objetivos con un alto nivel de recuperación, pero con baja precisión. La investigación de campo sugirió que los objetos no eran hornos de brea, sino una clase de características históricas diferente. Este estudio sugiere que los modelos entrenados con objetos simulados son un complemento útil a los kits de herramientas

© The Author(s), 2026. Published by Cambridge University Press on behalf of Society for American Archaeology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

de análisis lidar y pueden integrarse directamente en los flujos de trabajo de investigación arqueológica de campo.

Keywords: convolutional neural network (CNN); deep learning; lidar; procedural generation; tar kilns

Palabras clave: red neuronal convolucional (CNN); aprendizaje profundo; lidar; generación procedimental; hornos de brea

High-resolution digital surface datasets have become widely available to archaeologists over the last several decades. Archaeologists have responded by developing methods to locate archaeological sites in these data at a landscape scale. These methods include training deep learning (DL) models to automatically detect archaeological features.

Although rapidly locating archaeological sites at a landscape scale has implications for research and cultural resource management (CRM), limited training dataset availability can hinder large-scale model deployment. Training DL models on procedurally generated objects could eliminate this bottleneck. Procedural generation—algorithms that quickly generate content—can automate archaeological object creation, eliminating the need for time-consuming training dataset development. We used a structure type located during review of lidar-derived visualizations in the Kisatchie National Forest in Louisiana to test this approach. These structures look like historic tar kilns but have an unusual form and thus could not be detected using models trained on existing datasets. We implemented two procedural generation workflows for creating simulated structures and used these datasets to train DL models. We assessed these trained models' ability to detect the possible tar kilns and used the predictions to test our hypothesis about their function.

In this article, we use DL terminology to refer to key parts of our detection procedure. In archaeology, "feature" commonly refers to structural/nonportable components of archaeological sites. However, in DL contexts, the term "object" refers to detected items, whereas "feature" refers to an image/object's characteristics (e.g., an edge) that a DL model may use as part of its object recognition workflow. Therefore, in this study, *object* refers to the archaeological "features" we hope to detect. Other DL terms are defined at first use. A glossary is also included in Supplementary Material 1.

Background

Archaeological Deep Learning with Neural Networks

DL is a form of machine learning, an approach that uses algorithms to make predictions about datasets. DL models can be supervised, using information gained from training datasets to make more accurate predictions. Supervised machine learning is not new—for example, linear regression is a form of supervised machine learning—but "deep" learning models are unique because they use neural networks. Neural networks are models composed of many layers of interconnected "neurons" that theoretically mimic the structure of a human brain. As data are fed through a neural network's layers, specific features within the data activate specific neurons, which in turn activate neurons in the layers to which they are connected. This training procedure produces "weights," numbers that tell a neural network how to interpret new data. Neural networks can perform tasks like object recognition and classification that were historically difficult for computers, making them valuable for complex archaeological tasks (e.g., Fetaya et al. 2020; Navarro et al. 2022), including the aim of this study: locating archaeological sites in lidar data.

Archaeologists have used neural networks to successfully detect diverse archaeological object types across geographic settings; for example, agricultural terraces in Sāmoa (Quintus et al. 2023), buildings in the Maya region (Somrak et al. 2020), and barrows in the Netherlands (Verschoof-van der Vaart et al. 2020). Many recent lidar-focused archaeological DL implementations use some form of convolutional neural network (CNN; see Bonhage et al. 2021; Bundzel et al. 2020; Carter et al. 2021; Character et al. 2024; Küçükdemirci et al. 2023; Somrak et al. 2020; Suh and Ouimet 2023; Suh et al. 2021; Verschoof-van

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

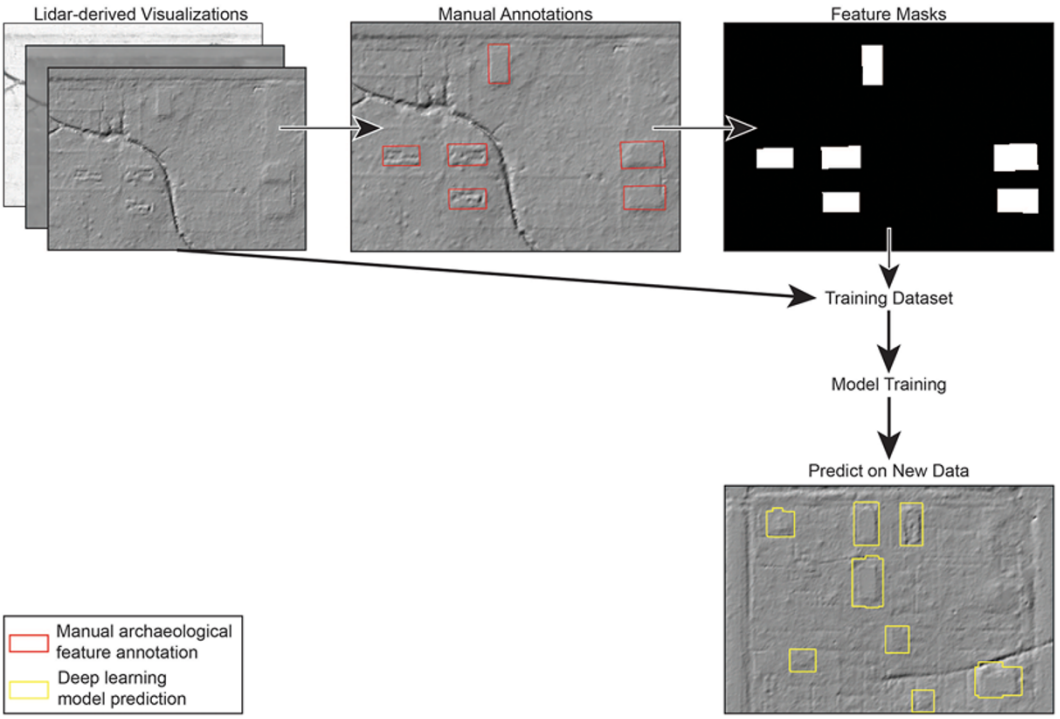


Figure 1. Idealized DL workflow with historic foundations as an example object.

der Vaart et al. 2023). This model structure uses filtering layers called convolution layers to emphasize image components (e.g., Karamitrou et al. 2022:Figure 1) and help the model learn which features characterize the object of interest. Therefore, CNNs are a good choice for detecting archaeological features in lidar data.

Training Datasets for Archaeological Deep Learning Models

Training a neural network to detect archaeological objects requires a training dataset of images of the object of interest (Figure 1, "lidar-derived visualizations") and complementary raster images indicating which parts of an image include an object and which parts do not (Figure 1, "feature masks"). When a neural network is shown a set of images capturing the variation within an object class—that is, the type of object—the trained model will have sufficient information to make a correct prediction when shown a new image of an object in that same class.

Comprehensive training dataset development requires a significant time investment. Archaeologists using DL for object detection often start by manually annotating—for example, drawing a shape around the archaeological object boundary—remote sensing data (Anttiroiko et al. 2023; Bonhage et al. 2021; Canedo et al. 2024; Carter et al. 2021; Fiorucci et al. 2022; Karamitrou et al. 2022; Somrak et al. 2020; Soroush et al. 2020; Verschoof-van der Vaart 2022; Verschoof-van der Vaart and Lambers 2019; Verschoof-van der Vaart et al. 2020) to create training datasets consisting of hundreds or thousands of "masked" archaeological object images. Other projects use existing archaeological site data from state/national cultural heritage databases or prior research projects (Banasiak et al. 2022; Casini et al. 2023; Quintus et al. 2023; Sobotkova et al. 2024; Suh and Ouimet 2023; Trier et al. 2021). However, even where databases of previously documented archaeological sites exist, transforming geospatial data into training data requires additional preparation (e.g., Casini et al. 2023). Further, certain archaeological objects may be so underdocumented or uncommon that large, comprehensive training datasets are impossible to create.

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

Small Training Datasets Complicate DL

In general, using large training datasets increases model performance, at least to a point (see Gravel-Miguel et al. 2025). Small training datasets can lead to overfitting, an outcome in which models learn to detect objects in the training datasets but cannot detect objects outside that dataset. One common approach to avoid overfitting is data augmentation, a series of image manipulations—for example, flipping, stretching, or blurring—applied to introduce variation in and increase the size of a training dataset. Data augmentation can improve performance on relatively small training datasets (e.g., Character et al. 2021; Somrak et al. 2020). However, this method is insufficient to capture the diversity required to train a well-performing model when it is applied on very small—fewer than 100—training datasets.

Some archaeologists use other approaches to creatively address the training dataset bottleneck; Meng (2023) cropped Neolithic rondel images and placed them into new topographic settings, Gallwey and colleagues (2019) used lunar surface craters as a training dataset for a historical mining pit detection model, and Vadineanu and colleagues (2024) proposed a rapid annotation workflow using a trained CNN to generate object masks. Outside archaeology, simulated sonar data are sometimes used for training underwater object detection models (e.g., Sung et al. 2020). To our knowledge, however, no one has yet attempted to train a DL model with a procedurally generated training dataset for terrestrial archaeological applications.

Procedural Generation

Procedural generation techniques create content, such as video game assets (see the discussion in Togelius et al. 2016). These techniques rely on different algorithms depending on the desired output content; in a video game context, for instance, a random number generator can create simple obstacles for a player to encounter (Brewer 2017:Figure 3), whereas noise generation algorithms can build extensive simulated landscapes (Parberry 2014; Smelik et al. 2009).

Togelius and colleagues (2016) note several characteristics of ideal procedural generation. Their concepts of speed, reliability, and controllability are especially relevant. Creating a procedurally generated training dataset should be quicker than manual annotation of a similar number of real archaeological objects. Further, the output dataset needs to be reliable (realistic enough) because an unrealistic object form or placement may introduce variation into the final model that is not present in real objects, increasing noise in its predictions. Finally, users need to control the generated data to tune the final training dataset to their needs. For instance, users might adjust the final simulated object size or size range so the output dataset matches the object of interest. A procedural generation algorithm with these characteristics could produce a sufficient DL model training dataset.

Case Study

Study Area

The Kisatchie National Forest (KNF) was established in 1930 on former logging company lands (Burns et al. 1994) and covers more than 600,000 acres (approximately 243,000 ha) in north-central Louisiana (Figure 2). Its five ranger districts are located across the Tertiary / Southern Tertiary Uplands ecoregion, characterized by hilly, dissected terrain with a mix of forest and wetland plant communities (Daigle et al. 2006). Decades of archaeological survey and excavation on the KNF show a long history of human occupation, from Paleoindian/Archaic use of the landscape's lithic resources to the dramatic landscape changes associated with the late nineteenth- and early twentieth-century lumber industries (Anderson and Smith 2003:15–20, 349–365, 437–451).

Unknown Archaeological Objects in the KNF

While examining a lidar dataset in the KNF's Catahoula Ranger District, we identified 12 structures in the vicinity of Camp Livingston, a World War II–era military installation, that are visually similar to historic tar kilns. Tar kilns are circular, earthen structures used to extract tar, pitch, and resin from pine trees for industrial and domestic uses. Commonly used from the eighteenth to the early twentieth centuries (Southerlin et al. 2021), tar kilns (Figure 3) consisted of stacked pine staves covered with earth

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

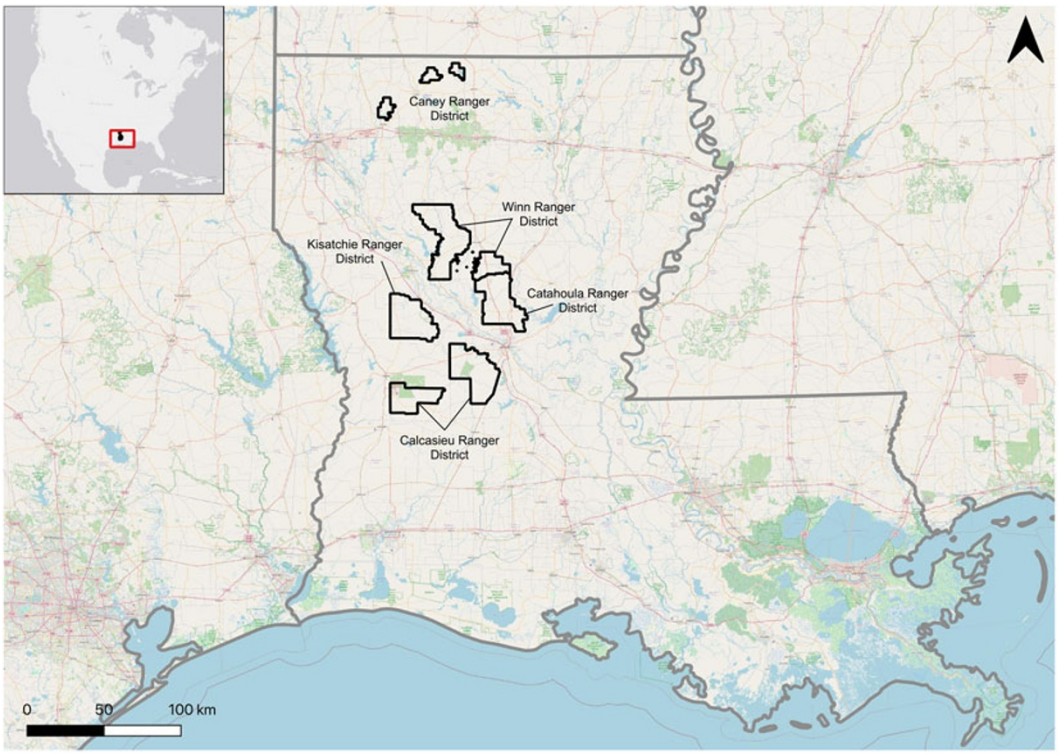


Figure 2. Map of KNF ranger districts, Louisiana (Basemaps: Open Street Map, ESRI).

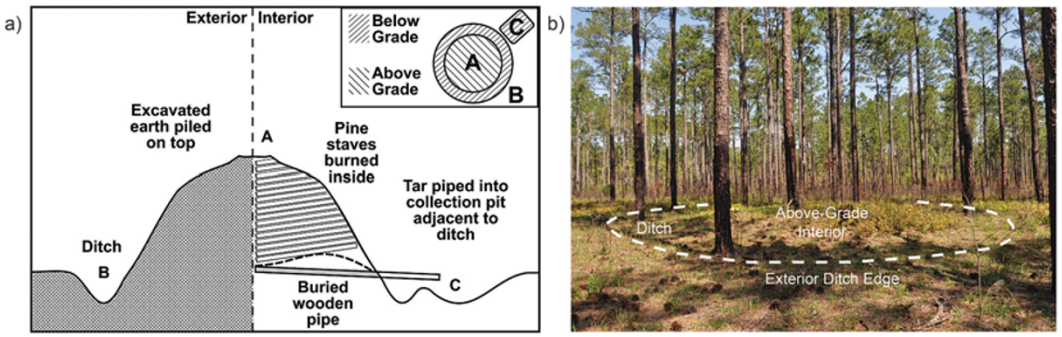


Figure 3. (a) Typical tar kiln cross section (based on Combes 1974); (b) archaeological tar kiln (Snitker et al. 2022).

excavated from a surrounding trench. The pine was burned in a low-oxygen environment to extract tar that was then piped to a barrel in a nearby collection pit. When located archaeologically, the only remaining tar kiln components are typically a below-grade ditch surrounding an above-grade interior and sometimes a visible collection pit on the trench exterior (Snitker et al. 2022).

The objects identified in the KNF are circular objects with an adjacent pit and thus share morphological attributes with documented tar kilns (Figure 4). However, these objects have an above-grade berm surrounding a below-grade interior (Figure 4, bottom), whereas elsewhere in the US Southeast (Figure 4, top) tar kilns have an above-grade interior surface surrounded by a ditch (Snitker et al. 2022). Tar kilns have not yet been documented archaeologically in Louisiana (Charles McGimsey, personal communication 2024), but they are referenced in historical accounts (Forbes 1921; Touchstone 1985) and have been documented elsewhere in the Southeast (Combes 1974; Hart 1986; Herbert et al. 2018;

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

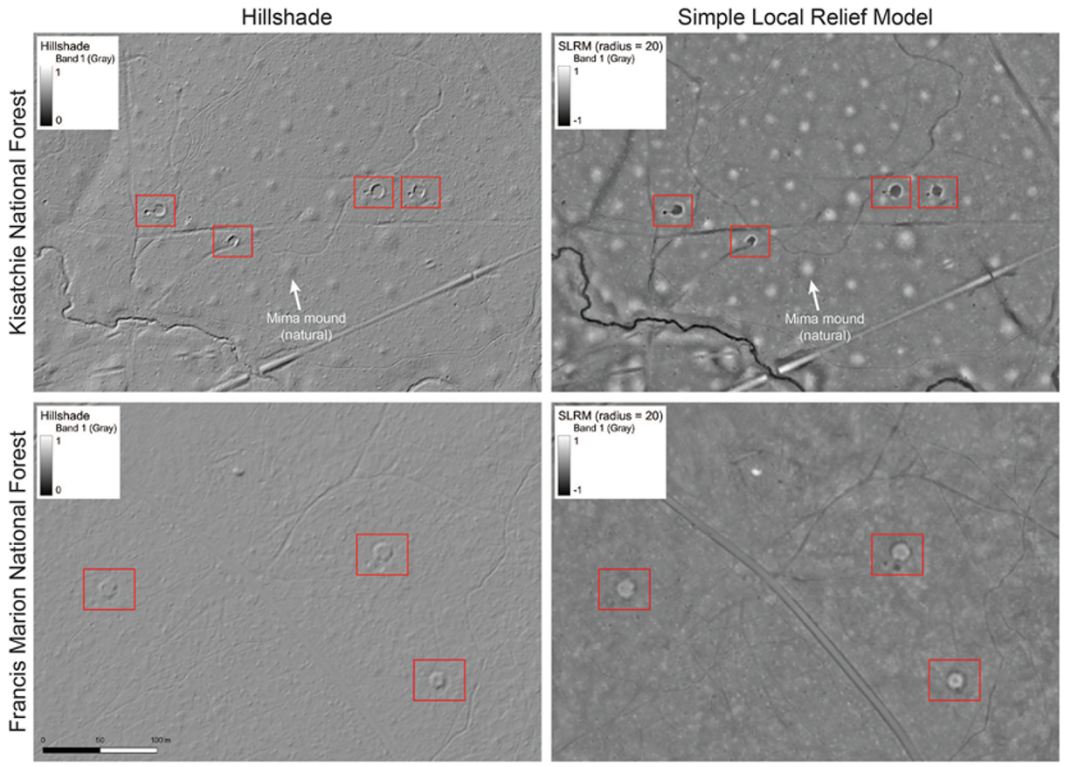


Figure 4. (*Top*) Unknown archaeological objects in the KNF, Louisiana, surrounded by mima mounds; (*bottom*) tar kilns in the FMNF, South Carolina.

Southerlin et al. 2021). Therefore, if the objects are tar kilns, they represent a new research and resource management opportunity.

Given the forest's size, automatic object detection with DL is an appropriate approach for locating other similar features. However, the small number of possible tar kilns (hereafter referred to as "targets") and their unique morphological characteristics complicate our ability to train and apply a DL model. Using data augmentation techniques alone on this small training dataset ($n = 12$) would create a model subject to overfitting. Although tar kilns are common enough in the Southeast that regional archaeological tar kiln datasets exist (Harrup 2013; Snitker et al. 2022; Southerlin et al. 2021), applying a model trained on southeastern tar kilns to the KNF does not work because of the targets' distinctive shape. Therefore, to train a model to detect objects like these targets, we need many more images of objects with the same form.

Data and Methods

Overview

We developed two methods for creating simulated training datasets and their associated annotation masks, both of which use procedural generation to modify real lidar-derived digital elevation models (DEMs; henceforth, Methods 1 and 2). Because of the targets' unique form, we did not have a training dataset of real objects against which to compare models trained on procedurally generated objects. As a baseline, we also trained a model on real tar kilns from the Francis Marion National Forest (FMNF) in South Carolina. To make this training dataset better match the targets, we inverted the DEM (henceforth, Method 3) before training. We then used these three datasets to train DL (Mask R-CNN) models with the same model parameters and training dataset size. To test the trained models' performance, we applied them to the section of the KNF containing the 12 targets. Model performance was assessed with common

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

machine learning metrics. We then used these predictions to ground-truth the objects in the field and test our hypothesis about the objects' function.

Lidar-Derived DEMs

For the Method 1 and 2 datasets, we downloaded and merged 0.5 m^2 resolution lidar-derived DEMs covering the entire KNF (US Geological Survey 2021) from the USGS National Map data downloader. We picked a ranger district within the KNF with similar topography to our prediction area, visually verified that it did not contain any objects resembling the targets, and selected a ~45,000-acre area (180 km^2) within the district (Method 1 DEM and Method 2 DEM; Figure 5). For the Method 3 training dataset, we used the NOAA Digital Coast Data Access Viewer to download a ~22,500 acre (90 km^2) 1 m^2 resolution DEM (Method 3 DEM; Figure 5) from the FMNF (OCM Partners 2017) containing 379 known tar kilns.

Model performance was assessed by applying the trained models to a lidar-derived DEM (Model Prediction DEM; Figure 5) covering the area where the targets were first located. We downloaded this 1,010-acre (approximately 4 km^2) Model Prediction DEM from the USGS National Map Data Downloader in the original project resolution of 0.5 m^2. This raster includes the 12 targets, as well as a small section of Camp Livingston. The area around Camp Livingston has many anthropogenic (e.g., historic structure foundations) and natural (e.g., mima mounds) objects, providing an opportunity to assess trained model performance in areas with a high number of potential false positives. To ensure all training datasets had similar attributes, we resampled all 0.5 m^2 resolution maps to a 1 m^2 resolution to match the dataset used for Method 3. The approach described later can be applied on higher-resolution lidar; however, 1 m^2 resolution lidar is accessible for much of the United States, allowing this workflow to be reproducible across different study areas.

Creating Simulated Objects

We developed three methods to create our training datasets:

1. **Method 1:** Create a training dataset comprising fully simulated targets (matching the objects depicted in Figure 4).
2. **Method 2:** Create a training dataset comprising "simple" targets (fully simulated circular berms without an adjacent pit).
3. **Method 3:** Invert an existing DEM featuring confirmed tar kilns ($n = 379$) to create a training dataset of "real" data.

Methods 1 and 2 are procedural generation approaches that create simulated training datasets. Python code for these methods can be found in the project code repository (see the Data Availability Statement) with detailed instructions on installation and usage. Method 3 uses real tar kiln visualizations and can be implemented in GIS software. These approaches are described in detail later. Specific parameters, such as original object sizes and procedurally generated object component size ranges, can be found in Supplementary Materials 2 and 3.

Method 1: Fully Simulated Tar Kilns

This workflow requires four inputs: a DEM, a vector drainage file covering the DEM, the desired number of objects to add to the DEM (n), and a desired output object radius (r). The script then adds that number of simulated objects to the DEM using the following procedure.

Create an Exclusion Model. First, the script creates an exclusion model to determine object placement. Tar kilns are not found in drainages or on very steep slopes (see, e.g., Snitker et al. 2022), nor are the documented targets. Therefore, our exclusion model (Figure 6a) removes areas of the DEM with high slope (>10%). The model also removes areas within drainages or within the vicinity of drainages.

Generate Centroids. Within the suitable areas identified by the exclusion model, the script selects n random points representing the future simulated target centroids. The script then removes any point within the r value provided by the user (Figure 6b) to avoid creating unrealistically overlapping simulated

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

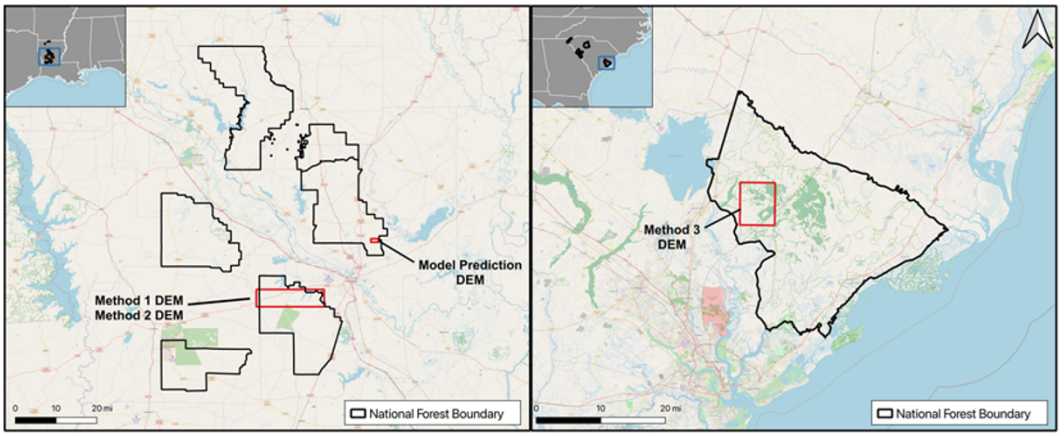


Figure 5. (*Left*) Training data for Methods 1 and 2 in the KNF; (*right*) training data for Method 3, FMNF (Basemap: Open Street Map).

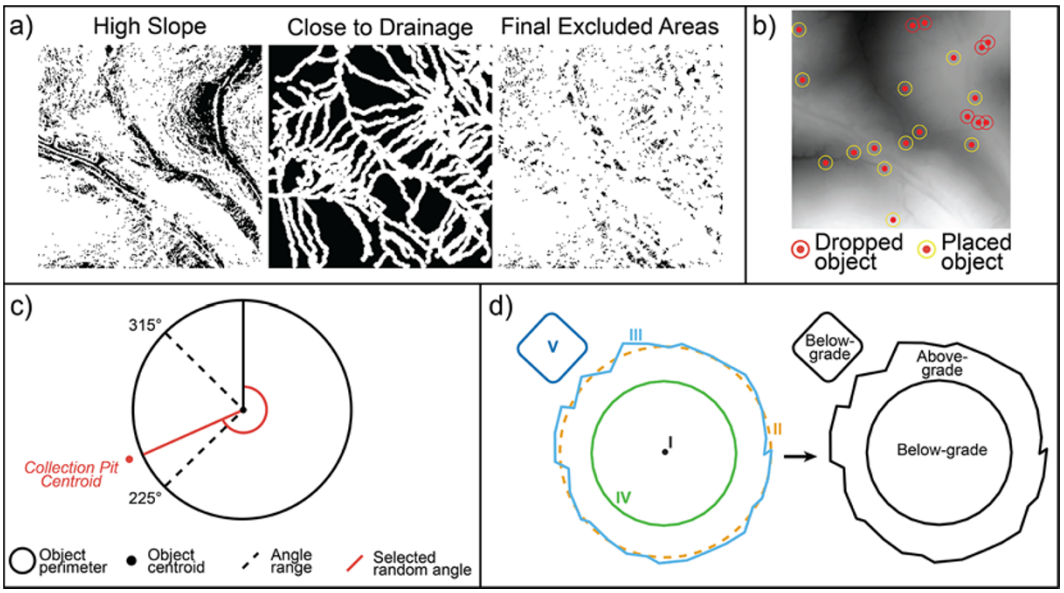


Figure 6. (a) Simulated object placement exclusion model on a 256 × 256-pixel tile (black areas represent suitable object placement locations); (b) simulated object placement in a 256 × 256-pixel tile; objects too close together are dropped; (c) collection pit placement rules; (d) the script generates a tar kiln perimeter (II) around each placed point (I) as a circular buffer, modifies it with random noise (III), and then generates the tar kiln interior (IV) and collection pit (V).

objects. If this process results in a final number of centroids less than *n*, the script selects a new set of random points until it finds a set of *n* points that do not overlap.

Create an Object around Each Centroid. Around each generated point ("I" in Figure 6d), the script creates several vector shapes. To form the tar kiln and the tar collection pit, the script creates a circular buffer around the generated point, defining the object's perimeter ("II" in Figure 6d). The script then generates random noise and applies it to the perimeter to create an imperfect exterior line, simulating postdepositional processes ("III" in Figure 6d). Then, the script creates a smaller circular buffer ("IV" in Figure 6d). These two shapes define the tar kiln's exterior berm ("III" in Figure 6d) and the below-grade interior's perimeter ("IV" in Figure 6d), respectively. The script then creates a collection pit next to each tar kiln by selecting a randomly offset pit centroid within a range of azimuths observed in the target

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

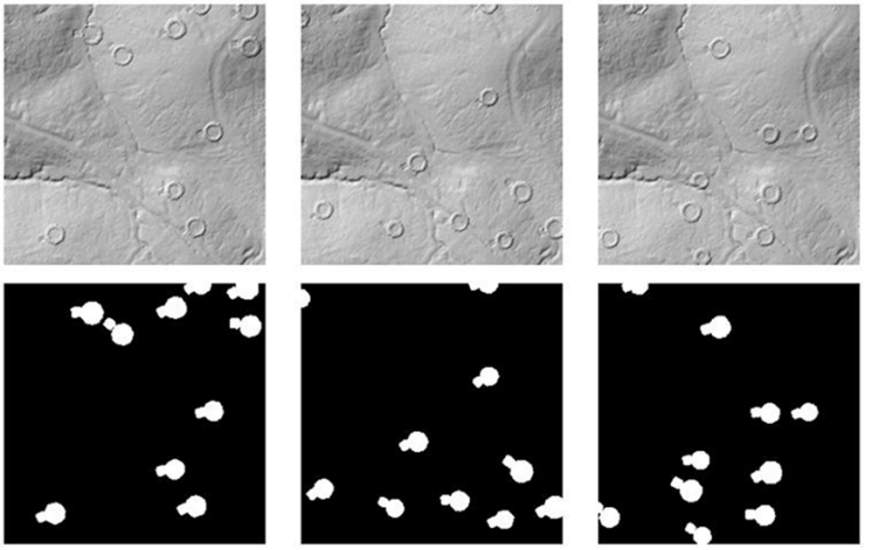


Figure 7. (*Top*) Different iterations of procedurally generated targets on the same 256 × 256-pixel tile; (*bottom*) their associated generated annotation masks.

objects (Supplementary Material 2:Table 1). These vector shapes are used to add or subtract elevation from the DEM to create the simulated object. The output is a real DEM in which simulated tar kilns have been embedded (Figure 7, top). Throughout this process, the script uses the object's vector shapes to create an annotation shapefile and turns that vector annotation into an annotation raster (Figure 7, bottom), eliminating the need for manual annotation.

To create this study's training dataset, we ran the script with $n = 379$, to match the training dataset used in Method 3. This allowed us to ascertain whether the performance of each trained model varied based on the type of training data (real vs. simulated) rather than the number of objects in the training dataset (see Table 1 for all parameters).

Method 2: Simple Tar Kilns

Method 1 followed the assumption that a simulated training dataset would be most effective when it closely matched the object of interest. To test this assumption, we also developed a workflow for "simple" objects: circular objects with an above-grade exterior berm and no adjacent pit.

This procedure uses Method 1's exclusion process to define object centroids. Then, using a user-provided minimum–maximum range, the script creates the berm and interior for each centroid following the Method 1 procedure. These vector shapes are then used to add and subtract height from the underlying DEM. As with Method 1, we ran the script with $n = 379$ (see Table 1).

Method 3: Inverted DEM

The FMNF contains hundreds of tar kilns that are characterized by a ditch surrounding an above-grade center (Snitker et al. 2022). Because the Kisatchie targets are "inverted" compared to the FMNF tar kilns, we hypothesized that inverting the lidar-derived DEM of FMNF tar kilns might be sufficient to train a target detection model. Therefore, we inverted the raster using a raster calculator tool (available in most GIS software), ensuring that the resulting raster retained its original range of values. The inverted tar kilns' profiles demonstrate that they more closely resemble the targets (Figure 8).

Model Training

Training Data. Training datasets consist of both image data and annotation masks that label each object. The Method 1 and 2 scripts automatically generate an annotation mask raster in which each object is

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

Table 1. Final Training Dataset Metrics for Methods 1 and 2.

Method	Metric	Parameter in Final Training Dataset
Method 1	Objects in final raster	379
	Input object radius	7 m (23 ft), creates objects between 10 and 15 m (32–49 ft)
	Stream buffer	10 m (32 ft)
	Slope threshold	10%
	Execution time	15.7 minutes
Method 2	Objects in final raster	379
	Object radius range (input)	3.0–3.5 m (10–16 ft), creates objects between 10 and 15 m (32–49 ft)
	Stream buffer	10 m (32 ft)
	Slope threshold	10%
	Execution time	5.83 min

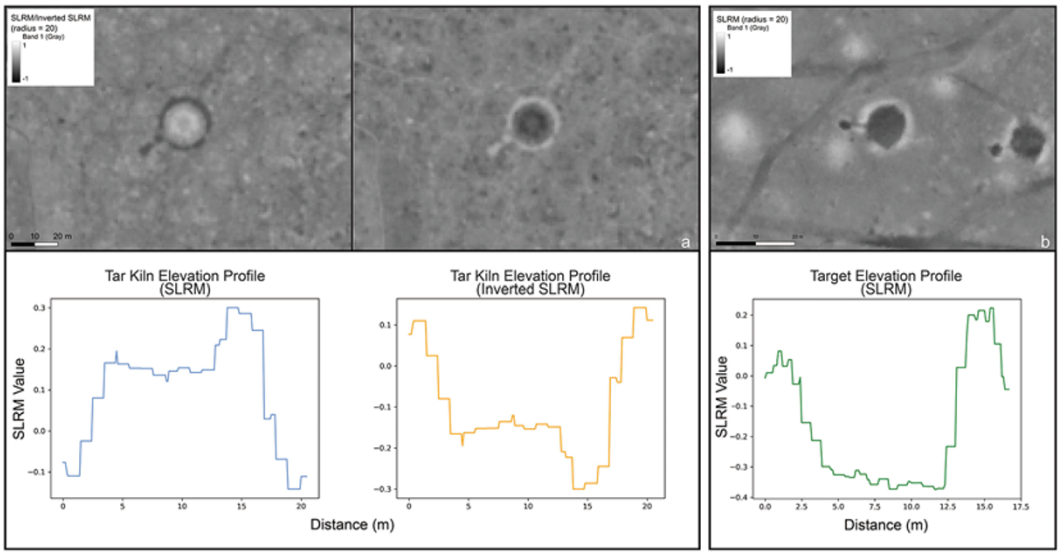


Figure 8. (a) SLRM (*top, left*) from the FMNF and inverted SLRM (*top, right*) with shared scale and legend, featuring a "standard" southeastern tar kiln at center. Also shown is the elevation profile of a tar kiln in SLRM (*bottom left*) and SLRM after DEM inversion (*bottom right*); (b) SLRM (*top*) and elevation profile (*bottom*) showing unmodified possible tar kilns from the KNF.

represented by pixels of a unique integer. For Method 3 we created a 20 m buffer around tar kiln centroids (identified by Snitker et al. 2022) that was wide enough to capture the tar kiln and collection pit, if present. We then converted these vector shapes to a raster, with each individual object represented by pixels of a unique integer.

To create the training images, we used the Relief Visualization Toolbox (Kokalj and Somrak 2019; Zakšek et al. 2011) to create a series of visualizations—hillshade, local dominance, negative openness, positive openness, slope, simple local relief model (SLRM) with a 20-pixel radius, and sky-view factor—for all DEMs. These visualizations were chosen based on prior work visualizing archaeological sites in lidar data (Guyot et al. 2021). For all methods, we tiled the visualizations and masks into 256 × 256 pixel tiles and scaled the pixel values in each image to values between 0 and 1 using code developed by Gravel-Miguel and colleagues (2025). This method creates a set of overlapping tiles (offset by 128 pixels) to double the number of training tiles and ensure that partial objects can still contribute to model training.

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

Model Development. All models were trained on an NVIDIA RTX A4000 GPU (16 GB of RAM) using PyTorch (Ansel et al. 2024) in a workflow and Mask R-CNN model implementation developed by our team. Mask R-CNN (He et al. 2017) is a DL model framework that can perform instance segmentation: when the model makes a prediction, it segments (i.e., divides) an image by giving a unique integer value to all the pixels belonging to a specific object within a class.

The model training workflow follows the one described in Gravel-Miguel and colleagues (2025), with slight modifications to fit the Mask R-CNN model requirements. The Python script first identifies appropriate training tiles (containing at least x pixels of the object of interest, based on a user-defined threshold). The script then separates the selected tiles into training, validation, and testing datasets (80%, 10%, and 10%, respectively). The script then preprocesses the tiles by combining three one-band visualizations into a three-band image by stacking them along a third dimension. Then, the script adds augmentations—for example, flipping, stretching, or blurring—to the training tiles and their respective masks; see Gravel-Miguel and colleagues (2025) for the list of augmentations applied along with their probabilities. It also computes the coordinates of a bounding box around each separate mask, a rectangular box that fully surrounds the object mask (required for Mask R-CNN model training). Training tiles, their masks, and bounding boxes are fed to the model in small batches until the model has seen them all and adjusted its neurons' weights accordingly at the end of one epoch (i.e., model training iteration) based on performance on the training dataset. The other datasets allow the user to assess the degree of overfitting (validation) and test the trained model's performance after all epochs (testing).

All models used "pretrained" weights; before training on this simulated archaeological dataset, the model backbone—the specific number and ordering of layers within a neural network; in this case, ResNet50—had already been pretrained on the ImageNet training dataset, a database of several million labeled images (Russakovsky et al. 2015). All models used a variable learning rate—how quickly the model adjusts its parameters after performance is calculated during an epoch—and the AdamW optimizer, the algorithm that helps the model change its parameters to reduce loss. All models were trained with a batch size—the number of images shown to the model at the same time—of eight.

Learning is assessed based on precision, recall, F1, and loss and is calculated after each epoch. These metrics are commonly used to assess DL model performance and are calculated based on the proportions of true positives, false positives, and false negatives. A true positive is a real object that the model correctly predicted. A false negative is an object that exists but that the model did not detect. A false positive is an object predicted by the model but that does not belong to the predicted class. Precision is the percentage of total predictions representing actual objects, whereas recall is the percentage of actual objects that are correctly predicted. A model can have high precision and low recall, meaning that although most objects it identified are real (i.e., few false positives), it also missed many real objects (i.e., many false negatives). In contrast, a model can have a high recall (i.e., detected most of the true positives) and low precision, meaning many of its predictions were incorrect (i.e., many false positives). F1 is the harmonic mean of precision and recall and is a common way of summarizing those metrics. During training, these metrics are calculated pixel by pixel: each pixel in the prediction mask that overlaps with a real object's pixel in the annotation mask is a true positive.

Loss is the final metric used to evaluate model performance during training. Loss describes the average inaccuracy of predictions during model training. Therefore, it is positively correlated with the number of incorrect predictions. After each epoch, the model weights are adjusted to minimize loss. Loss can be calculated using different functions; in this study, we use the procedure hard-coded into the PyTorch Mask R-CNN implementation (Ansel et al. 2024), which follows the procedure in the original Mask R-CNN publication (He et al. 2017) to calculate loss after each training epoch. That value is a sum of five different loss functions applied on the mask, the bounding box, and the label assigned to each prediction. All models were trained for 15 epochs, at which point metrics plateaued.

We primarily tested different visualization combinations across model trainings. To find the visualization parameters that created the best model, we followed an iterative process. Starting with the Method

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

3 data (the inverted DEM model), we trained a model with each unique combination of available visualizations. We then determined which model had the best F1 when calculated on the model's testing dataset. We found that a model trained on all tiles visualized to show local dominance produced the best metrics. All subsequent models were then trained on local dominance tiles.

Postprocessing

We developed a postprocessing procedure to clean model predictions after observing common false positives in initial model outputs. Most of the observed false positives were mima mounds, a natural mound prevalent in low-relief areas of the KNF (an example is labeled in Figure 4) and elsewhere in North America (see, e.g., discussion in Washburn 1988). Although both are circular, the elevation profile of a mima mound is distinct from the profile of our targets; the former are dome-like (convex), whereas the latter are pit-like (concave). Their profiles reflect this difference in form.

To remove mima mounds from the models' predictions, the postprocessing script calculated the major and minor axes for each object and sampled the underlying DEM every 10 cm, producing two profiles for each object (Figure 9, top). The script then used a linear regression to smooth the profiles and calculated their second derivatives (Figure 9, bottom), which indicate the direction of the curve. A profile with a negative second derivative indicates a dome-like object that is likely not a target. Conversely, a function with a positive second derivative indicates a pit-like object that is more likely to be a target. Using this information, the script classified all objects based on the sign of both of their profiles' second derivatives: an object with two positive second derivatives was classified as a likely target, an object with two negative second derivatives was classified as a likely mima mound, and an object with one positive and one negative second derivative was classified as ambiguous. For this project, we used this procedure to classify all objects predicted by each model and to remove likely mima mounds from the prediction dataset.

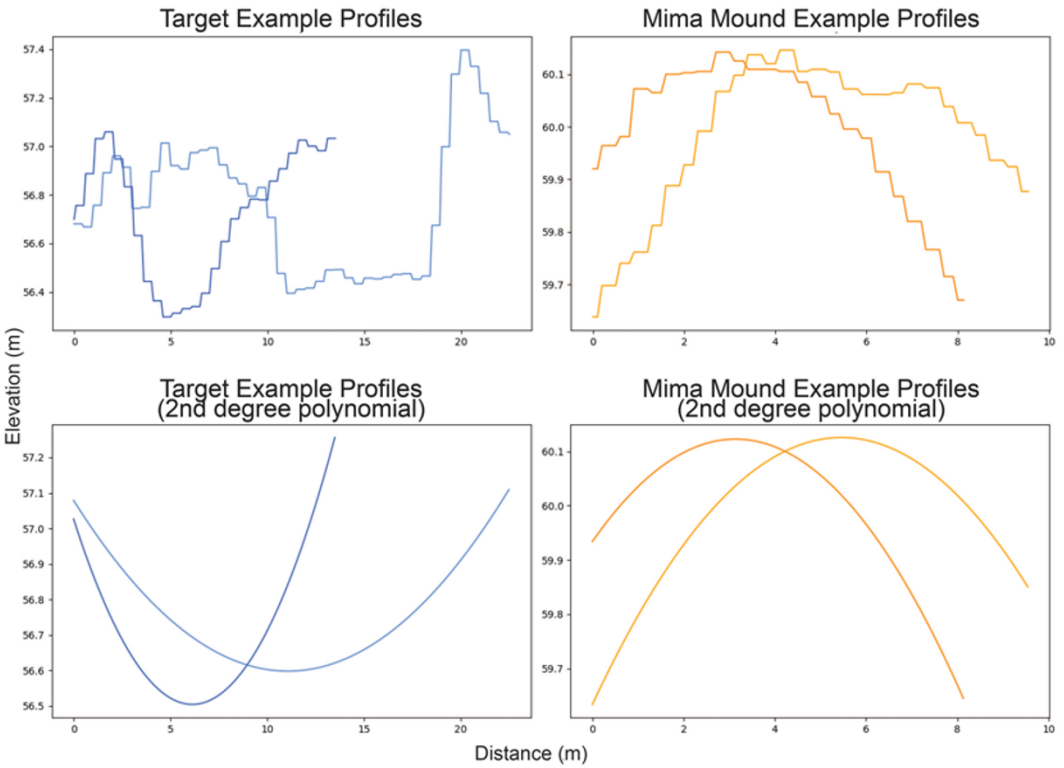


Figure 9. Perpendicular profiles from a single target (*left*) and mima mound (*right*) before (*top*) and after (*bottom*) smoothing.

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

To further improve prediction denoising, we dropped all very small objects (less than 30 m^2) from the dataset. Finally, because many of the observed false positives were drainage meanders, we also dropped all objects that intersected, were within 15 m of a drainage (determined using a stream extraction procedure on a 10 m resolution DEM with an initiation threshold of 1,000 pixels), or both. These denoising procedures were applied to all model predictions.

After postprocessing, we systematically examined the hillshaded lidar-derived DEM and labeled each remaining object based on its likely function. Supplementary Material 4 records the distribution of predictions across each class.

Ground-Truthing

After systematic review, we merged each unique "possible" target into a GeoPackage with the original 12 targets. We assigned each new possible target a confidence rating based on its appearance and proximity to the original targets. Eleven of the 12 original targets were visited in the field, photographed, and visually evaluated for their potential as historic tar kilns. A subset of the original targets ($n = 2$) was measured, mapped, and augered to evaluate for subsurface attributes unique to tar kilns such as charcoal-rich sediments, charred pine billets, a buried wooden tar pipe, or a hard-pan clay floor. A subsample of "high confidence" targets ($n = 2$) was also visited to evaluate their potential as tar kilns. Access to all targets was not possible because of the dense understory vegetation. The remaining unvisited targets were evaluated remotely using lidar-derived visualizations and satellite imagery in consultation with archaeologists from the KNF Heritage Program.

Results

Prediction Methods

Method 1: Procedurally Generated Tar Kilns. The simulated tar kiln model predicted 183 objects in the Model Prediction DEM. Postprocessing automatically removed 41 objects, leaving 142 predictions to assess manually. This model (Supplementary Material 4:Figure 1) located nine of the 12 targets and six additional objects that, on review, were close enough in appearance to the targets to warrant future ground-truthing.

The remaining objects ($n = 127$) did not appear to be tar kilns. These false positives included a range of natural and anthropogenic objects (Figure 10). The majority ($n = 91$) of the false positives were reservoirs (e.g., stock tanks/ponds) that have a similar form to the targets.

Method 2: Procedurally Generated "Simple" Tar Kilns. The Method 2 model predicted 2,032 objects in the Model Prediction DEM. Postprocessing removed 1,323 objects, leaving 709 objects for systematic review. This model (Supplementary Material 4:Figure 2) located all 12 targets and 11 additional objects warranting future ground-truthing. The remaining 686 objects were all false positives, most commonly drainages ($n = 132$), reservoirs ($n = 111$), or indistinct objects that could not be assigned to a specific class ($n = 123$).

Method 3: Inverted DEM. The Method 3 model predicted 1,858 objects in the Model Prediction DEM. Postprocessing removed 1,482 objects, leaving 376 objects for systematic review. This model (Supplementary Material 4:Figure 3) located 11 out of the 12 targets and seven additional objects warranting future ground-truthing. In total, this model had 369 false positives, primarily reservoirs ($n = 108$) and drainages ($n = 67$).

Ground-Truthing

In the field, the targets (Supplementary Material 4:Figure 4) appeared similar to tar kilns. However, auger tests in two of the original 12 targets revealed no concentrations of charcoal, billets, or wooden tar pipes nor a hard-pan clay floor. Natural undisturbed clays were encountered at approximately 30 cm (12 inches) below the surface of each potential kiln. All overlying deposits were interpreted as recently accumulated fill on top of the undisturbed clays. Given the proximity of these objects to a historic military base, combined with the absence of all diagnostic attributes of a tar kiln, we concluded that they are most

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

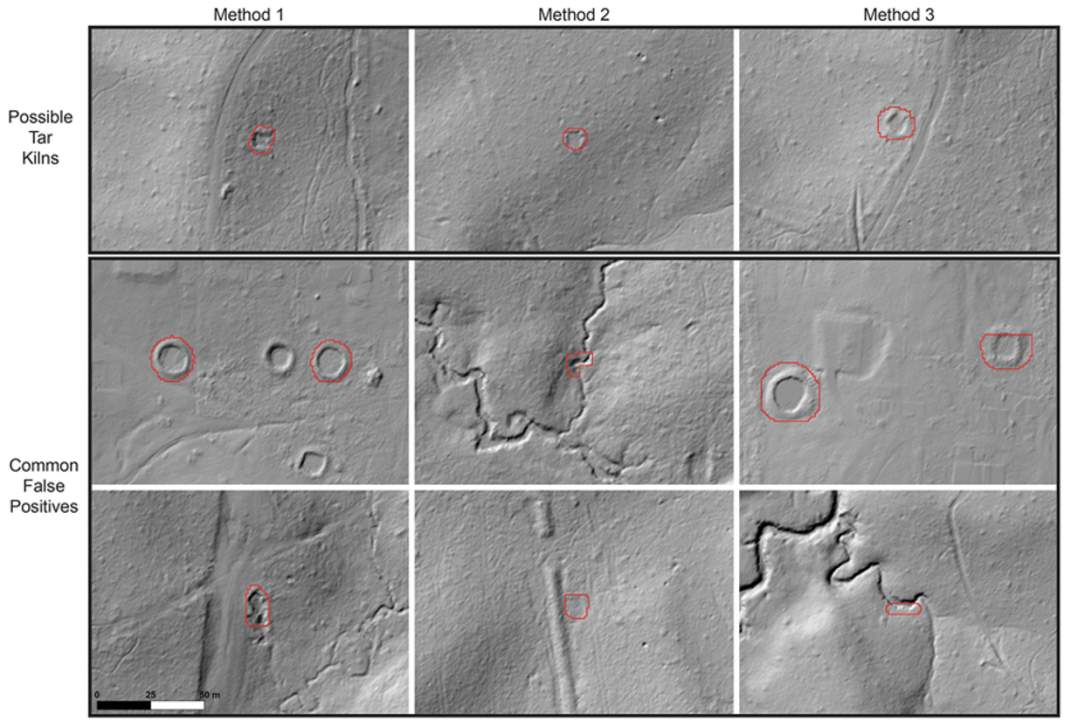


Figure 10. (*Top*) New possible targets located by all methods; (*bottom*) the two most common false-positive categories across each method. All images share the same scale.

likely associated with military training exercises. Review of a World War II–era military training manual (United States War Department 1944) suggests the features may be howitzer emplacements (Figure 11). Additional historical research and comparison with other, similar sites will be needed to draw a firm conclusion about object function. Based on these field observations and interpretations, we determined that all remaining unvisited "high confidence" targets near the historic military base were not tar kilns (Figure 10, top).

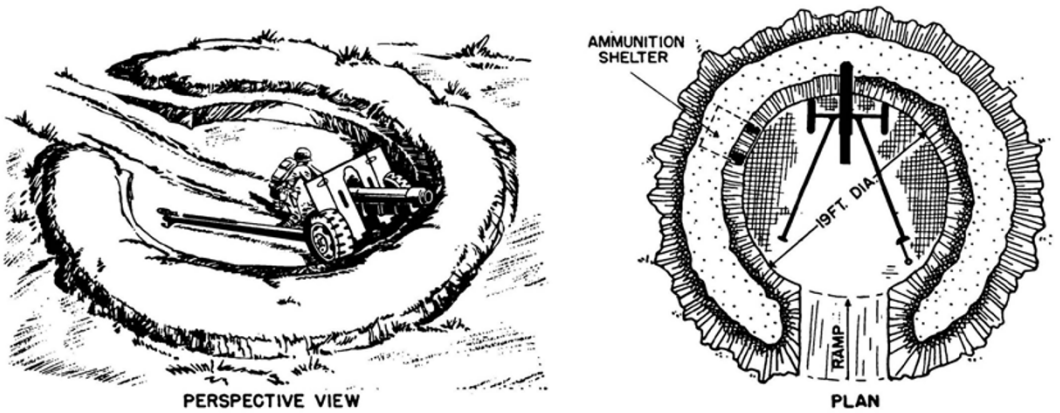


Figure 11. Howitzer emplacement (illustration modified from Figure 38 in Corps of Engineers Field Fortifications manual FM 5-15; United States War Department 1944:79).

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

Table 2. Object-by-Object Metrics for All Models When Applied to the Testing Area.

Method	Training Dataset	Recall	Precision	F1	True Positives	False Negatives	False Positives
1	Simulated tar kilns	0.750	0.063	0.117	9	3	133
2	Simple tar kilns	1.000	0.017	0.033	12	0	697
3	Inverted DEM	0.917	0.025	0.048	11	1	431

Notes: Metrics are calculated against the 12 original targets. The best metrics in each column are underlined.

Discussion

Model Performance

Comparing all methods to each other (Table 2) based on their F1 scores, the simulated tar kilns model (Method 1) performed the best, and the "simple" tar kilns model (Method 2) performed the worst. Although Method 1 failed to locate three of the targets, it had substantially fewer false positives overall, meaning its predictions were technically the most precise. However, despite Method 1's higher F1 score, Method 2 had the highest recall. All models located additional possible targets, a total of 12 unique objects across all models (Figure 10).

These results suggest that a procedurally generated training dataset can train a neural network to detect archaeological objects. Depending on the performance metric that the user wishes to maximize (precision vs. recall), a model trained in this way can perform better than a model trained on modified real data. The difference between the Method 1 and 2 metrics suggests that the more closely the procedurally generated object matches the target object (i.e., Method 1), the higher the precision. However, a training dataset comprising more "general" procedurally generated objects (i.e., Method 2) will increase recall.

Deep Learning in Archaeological Workflows

DL predictions on lidar-derived DEMs have the potential to contribute to CRM and archaeological research. Lidar data alone enable the identification of subtle archaeological landscape modifications, such as tar kilns, paths, tramways, homesteads, and roads in heavily forested environments, amplifying pedestrian archaeological surveys in which these features are often not visible to the naked eye. Incorporating DL predictions into visual lidar prospection can reduce the mental load and time associated with locating sites in these data. Resource managers can then use these predictions to develop targeted survey strategies and predictive models within land management areas. As land management shifts toward larger landscape-level restoration, having high-resolution, landscape-level archaeological data may feed directly into these project goals from an integrated resource management perspective. Lidar analysis is also frequently used in landscape archaeology research in the US Southeast (e.g., Davis et al. 2020, 2021; Roberts Thompson and Finch 2023), and DL predictions could aid in understanding site relationships at a landscape scale.

To use DL predictions in these management and research goals, reliable trained models must be available for a range of archaeological objects. This study demonstrates that archaeologists can use procedural generation to create datasets, ready for model training, of rare objects. These datasets are generated in minutes, substantially reducing the workload needed to train a model to generate predictions. Reviewed predictions can then streamline pre-field research, planning, and fieldwork. Although these predictions alone cannot contribute to assessment of eligibility for the National Register of Historic Places, which is key in the United States for evaluating adverse impacts of federal undertakings in a project area, they could help define areas of potential effect or target subsurface sampling before conducting Phase I or Phase II archaeological investigations. In a research project where a 100% inventory may be less important than locating specific site types, training models on procedurally generated datasets could produce predictions that allow researchers to target field reconnaissance, as was the case in this study. Here, we used our reviewed predictions to target field investigations in a limited geographic area and on a smaller number of objects. This fieldwork then allowed us to reject our hypothesis that these objects are historic tar kilns. Future fieldwork and review of the historical record could then inform conclusions on object

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

function and contribute to a broader understanding of military landscape use in Camp Livingston and the surrounding area.

Although these model predictions were demonstrably useful, model metrics are lower than other recent archaeological DL case studies (Table 2; see Verschoof-van der Vaart et al. [2023] for results on a similar object type). However, we suggest that low-precision models may still be valuable detection tools. For our case study (and other projects attempting to locate objects with a unique form in a new study area), a high-recall and low-precision model like Method 2 is acceptable; our team found it straightforward to remove/label most model false positives and create the final fieldwork dataset. However, high false-positive rates have implications for integrating this approach into a management or research workflow. Because the outputs require manual cleaning, some effort is shifted from field survey to pre-field preparation. Although we found that the effort/time commitment needed to clean DL predictions was less than that required to complete a pedestrian survey of the same area, further model development would likely improve detection precision. Here, all models were trained on what is still a relatively small training dataset ($<$400) to ensure comparable results across all methods. This same procedural generation workflow could create a much larger (thousands of objects) training dataset in a short time, likely increasing performance.

Along the same lines, this study demonstrates the value of postprocessing and training dataset manipulation in increasing DL model reliability. If false positives can be excluded through automatic denoising, manually cleaning outputs will take less time. The Method 2 and 3 outputs, for instance, included many ($>$1,500) false positives; however, we were able to quickly and effectively exclude those with a predictable form (mima mounds) using automated postprocessing. We believe this approach could be useful for other automated archaeological object detection workflows. Further, although Model 3 generally performed worse than the models trained on procedurally generated data, it still located 11 of 12 targets. Even though it is not a procedural generation approach, this result suggests that modifying existing training data may be another useful approach for overcoming the training dataset bottleneck.

Finally, our approach focused on a specific object type. Given the success of this approach for identifying and ground-truthing these historic military objects, we suggest this method could be applicable for other archaeological object types, such as historic infrastructure, visible in lidar-derived DEMs. However, the procedural generation approach described here would need to be modified according to the attributes of the object of interest. Building foundations, for instance, could also be simulated by selecting object locations with an exclusion model and building a geometric shape around those points, whereas reticulate objects like railroad grades require a different approach. For those objects, researchers could draw on existing procedural generation algorithms—for example, for road networks, as discussed by Kelly and McCabe (2006)—to embed realistic networks in lidar-derived DEMs.

Conclusion

We developed two methods for creating simulated archaeological objects and assessed their efficacy for training DL models compared to modified real archaeological object data. We found that our models trained on fully simulated archaeological training datasets could detect the targets, although metrics varied between the two models. Our DL predictions increased fieldwork efficiency and allowed us to determine that our targets had a different function than initially hypothesized. Therefore, we suggest that a simulated object model could be a useful "first pass" before conducting more in-depth archaeological object annotation or archaeological survey in areas where structural archaeological sites are expected. Given the successful application for this object type, procedural generation could likely be applied to other archaeological objects visible in lidar data.

Automated archaeological object detection has implications for CRM and archaeological research at local and national scales. However, training dataset availability remains a bottleneck for incorporating archaeological DL at these scales. The creative use of procedurally generated objects represents one way to expand our ability to quickly locate archaeological objects in lidar-derived visualizations, providing useful data for archaeological survey and testing.

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

Acknowledgments. We would like to thank John Mayer, Velicia Bergstrom, Steve Treloar, and Lisa Lewis of the Kisatchie National Forest, as well as members of the US Fish and Wildlife Service aerial lidar data collection team, for their support and subject matter expertise during the duration of this project.

Funding Statement. This material is based on work supported by the US Department of Agriculture, Forest Service cooperative agreement #22-PA-11080600-228. Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the USDA Forest Service. This institution is an equal opportunity provider.

Data Availability Statement. Lidar data are available through the USGS National Map Data Downloader (https://apps.nationalmap.gov/downloader/) and the NOAA Data Access Viewer (https://coast.noaa.gov/dataviewer/). All code used in this study is available in a public GitHub repository (https://github.com/NMC-CRS/simulated-training-data-for-archaeological-site-detection) or via the following DOI: https://doi.org/10.5281/zenodo.17082763.

Competing Interests. The authors declare none.

Supplementary Material. The supplementary material for this article can be found at https://doi.org/10.1017/aap.2025.10130.
Supplementary Material 1. Glossary of Deep Learning Terms (text and table).
Supplementary Material 2. Target Measurements (text and table).
Supplementary Material 3. Method 1–3 Parameters (text).
Supplementary Material 4. Systematic Review of Model Predictions (text, table, and figures).

References Cited

Anderson, David G., and Steven D. Smith. 2003. *Archaeology, History, and Predictive Modeling: Research at Fort Polk 1972–2002*, University of Alabama Press, Tuscaloosa.

Ansel, Jason, Edward Yang, Horace He, Natalia Gimelshein, Animesh Jain, Michael Voznesensky, Bin Bao, et al. 2024. PyTorch 2: Faster Machine Learning through Dynamic Python Bytecode Transformation and Graph Compilation. In *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, Vol. 2, pp. 929–947. Association for Computing Machinery, New York. https://doi.org/10.1145/3620665.3640366.

Anttiroiko, Niko, Floris Jan Groesz, Janne Ikäheimo, Aleksi Kelloniemi, Risto Nurmi, Stian Rostad, and Oula Seitsonen. 2023. Detecting the Archaeological Traces of Tar Production Kilns in the Northern Boreal Forests Based on Airborne Laser Scanning and Deep Learning. *Remote Sensing* 15(7):1799. https://doi.org/10.3390/rs15071799.

Banasiak, Paweł, Piotr Berezowski, Rafał Zapłata, Miłosz Mielcarek, Konrad Duraj, and Krzysztof Stereńczak. 2022. Semantic Segmentation (U-Net) of Archaeological Features in Airborne Laser Scanning—Example of the Białowieża Forest. *Remote Sensing* 14(4):995. https://doi.org/10.3390/rs14040995.

Bonhage, Alexander, Mahmoud Eltaher, Thomas Raab, Michael Breuß, Alexandra Raab, and Anna Schneider. 2021. A Modified Mask Region-Based Convolutional Neural Network Approach for the Automated Detection of Archaeological Sites on High-Resolution Light Detection and Ranging-Derived Digital Elevation Models in the North German Lowland. *Archaeological Prospection* 28(2):177–186.

Brewer, Nathan. 2017. Computerized Dungeons and Randomly Generated Worlds: From Rogue to Minecraft. *Proceedings of the IEEE* 105(5):970–977.

Bundzel, Marek, Miroslav Jaščur, Milan Kováč, Tibor Lieskovský, Peter Sinčák, and Tomáš Tkáčik. 2020. Semantic Segmentation of Airborne LiDAR Data in Maya Archaeology. *Remote Sensing* 12(22):3685. https://doi.org/10.3390/rs12223685.

Burns, Anna C., Ronald W. Couch, and United States Forest Service Southern Region. 1994. *A History of the Kisatchie National Forest*. US Dept. of Agriculture, Forest Service, Southern Region, Kisatchie National Forest, Pineville, Louisiana.

Canedo, Daniel, João Fonte, Rita Dias, Tiago do Pereiro, Luís Gonçalves-Seco, Marta Vázquez, Petia Georgieva, and António J. R. Neves. 2024. Automated Detection of Hillforts in Remote Sensing Imagery with Deep Multimodal Segmentation. *Archaeological Prospection* 32(2):297–311. https://doi.org/10.1002/arp.1958.

Carter, Benjamin P., Jeff H. Blackadar, and Weston L. A. Conner. 2021. When Computers Dream of Charcoal: Using Deep Learning, Open Tools, and Open Data to Identify Relict Charcoal Hearths in and around State Game Lands in Pennsylvania. *Advances in Archaeological Practice* 9(4):257–271. https://doi.org/10.1017/aap.2021.17.

Casini, Luca, Nicolò Marchetti, Andrea Montanucci, Valentina Orrù, and Marco Roccetti. 2023. A Human–AI Collaboration Workflow for Archaeological Sites Detection. *Scientific Reports* 13(1):8699.

Character, Leila, Tim Beach, Takeshi Inomata, Thomas G. Garrison, Sheryl Luzzadder-Beach, J. Dennis Baldwin, Rafael Cambranes, Flory Pinzón, and José L Ranchos. 2024. Broadscale Deep Learning Model for Archaeological Feature Detection across the Maya Area. *Journal of Archaeological Science* 169:106022. https://doi.org/10.1016/j.jas.2024.106022.

Character, Leila, Agustin Ortiz Jr., Tim Beach, and Sheryl Luzzadder-Beach. 2021. Archaeologic Machine Learning for Shipwreck Detection Using Lidar and Sonar. *Remote Sensing* 13(9):1759.

Combes, John D. 1974. Charcoal Kilns and Cemetery at Paris Mountain State Park. *Institute of Archeology and Anthropology Notebook* 6(1):3–17.

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

Daigle, Jerry J., Glen E. Griffith, James M. Omernik, Patricia L. Faulkner, Richard P. McCulloch, Lawrence R. Handley, Latimore M. Smith, and Shannen S. Chapman. 2006. *Ecoregions of Louisiana (Color Poster with Map, Descriptive Text, Summary, Tables, and Photographs). 1:1,000,000*. US Geological Survey, Reston, Virginia.

Davis, Dylan S., Gino Caspari, Carl P. Lipo, and Matthew C. Sanger. 2021. Deep Learning Reveals Extent of Archaic Native American Shell-Ring Building Practices. *Journal of Archaeological Science* 132:105433. https://doi.org/10.1016/j.jas.2021.105433.

Davis, Dylan S., Robert J. DiNapoli, Matthew C. Sanger, and Carl P. Lipo. 2020. The Integration of Lidar and Legacy Datasets Provides Improved Explanations for the Spatial Patterning of Shell Rings in the American Southeast. *Advances in Archaeological Practice* 8(4):361–375.

Fetaya, Ethan, Yonatan Lifshitz, Elad Aaron, and Shai Gordin. 2020. Restoration of Fragmentary Babylonian Texts Using Recurrent Neural Networks. *PNAS* 117(37):22743–22751.

Fiorucci, Marco, Wouter B. Verschoof-van der Vaart, Paolo Soleni, Bertrand Le Saux, and Arianna Traviglia. 2022. Deep Learning for Archaeological Object Detection on LiDAR: New Evaluation Measures and Insights. *Remote Sensing* 14(7):1694. https://doi.org/10.3390/rs14071694.

Forbes, R. D. 1921. The Future of the Naval Stores Industry and Pine Forests in Louisiana: With Lessons Applicable to Other Naval Stores States. In *Naval Stores: History, Production, Distribution, and Consumption*, edited by Thomas Gamble, pp. 240–244. Review Publishing and Printing Co., Savannah, Georgia.

Gallwey, Jane, Matthew Eyre, Metthew Tonkins, and John Coggan. 2019. Bringing Lunar LiDAR back down to Earth: Mapping Our Industrial Heritage through Deep Transfer Learning. *Remote Sensing* 11(17). https://doi.org/10.3390/rs11171994.

Gravel-Miguel, Claudine, Grant Snitker, Jayde N. Hirniak, Katherine Peck, and Alex Fetterhoff. 2025. Semantic Segmentation of Archaeological Features on Public Lands: Case Study of Historical Cotton Terraces within the Piedmont National Wildlife Refuge, Georgia, USA. *Advances in Archaeological Practice* 13(2):291–314. https://doi.org/10.1017/aap.2025.1.

Guyot, Alexandre, Marc Lennon, and Laurence Hubert-Moy. 2021. Objective Comparison of Relief Visualization Techniques with Deep CNN for Archaeology. *Journal of Archaeological Science: Reports* 38:103027. https://doi.org/10.1016/j.jasrep.2021.103027.

Harrup, Matthew J. 2013. Tar Kilns of Goose Creek State Park: History and Preservation. Master's thesis, Department of History, East Carolina University, Greenville, North Carolina.

Hart, Linda P. 1986. *Excavations at the Limerick Tar Kiln Site (38BK472)*. Cultural Resource Management Report 86-52. Report on file at the Francis Marion and Sumter National Forest Supervisor's Office, Columbia, South Carolina.

He, Kaiming, Georgia Gkioxari, Piotr Dollár, Ross Girshick. 2017. Mask R-CNN. *arXiv*:1703.06870. https://doi.org/10.48550/arXiv.1703.06870.

Herbert, Joseph M., Linda F. Carnes-McNaughton, William J. Feltz, Michelle Hagstrom Parsons, and Jonathan Schleier. 2018. Anatomy of a Tar Kiln. *North Carolina Archaeology* 67:48–76.

Karamitrou, Alexandra, Fraser Sturt, Petros Bogiatzis, and David Beresford-Jones. 2022. Towards the Use of Artificial Intelligence Deep Learning Networks for Detection of Archaeological Sites. *Surface Topography: Metrology and Properties* 10(4):044001.

Kelly, George, and Hugh McCabe. 2006. A Survey of Procedural Techniques for City Generation. *ITB Journal* 7(2):5.

Kokalj, Žiga, and Maja Somrak. 2019. Why Not a Single Image? Combining Visualizations to Facilitate Fieldwork and On-Screen Mapping. *Remote Sensing* 11(7):747.

Küçükdemirci, Melda, Giacomo Landeschi, Mattias Ohlsson, and Nicolo Dell'Unto. 2023. Investigating Ancient Agricultural Field Systems in Sweden from Airborne LIDAR Data by Using Convolutional Neural Network. *Archaeological Prospection* 30(2):209–219. https://doi.org/10.1002/arp.1886.

Meng, Yuqi. 2023. Identification of Neolithic Circular Enclosures through Aerial Imagery: A Study about Pattern Recognition and Deep Learning Techniques. Master's thesis, Department of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands.

Navarro, Pablo, Celia Cintas, Manuel Lucena, José Manuel Fuertes, Rafael Segura, Claudio Delrieux, and Rolando González-José. 2022. Reconstruction of Iberian Ceramic Potteries Using Generative Adversarial Networks. *Scientific Reports* 12(1):10644.

OCM Partners. 2017. *2017 SC DNR Lidar DEM: Coastal Counties*. NOAA National Centers for Environmental Information. https://www.fisheries.noaa.gov/inport/item/57112, accessed August 19, 2024.

Parberry, Ian. 2014. Designer Worlds: Procedural Generation of Infinite Terrain from Real-World Elevation Data. *Journal of Computer Graphics Techniques* 3(1):74–85. https://www.jcgt.org/published/0003/01/04/paper.pdf, accessed February 3, 2026.

Quintus, Seth, Dylan S. Davis, and Ethan E. Cochrane. 2023. Evaluating Mask R-CNN Models to Extract Terracing across Oceanic High Islands: A Case Study from Sāmoa. *Archaeological Prospection* 30(4):477–492. https://doi.org/10.1002/arp.1909.

Roberts Thompson, Amanda D., and Jonathan Finch. 2023. Reconstructing the Layout of a Coastal Georgia Plantation: Applications of LiDAR. *Journal of Archaeological Science: Reports* 47:103798. https://doi.org/10.1016/j.jasrep.2022.103798.

Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, et al. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 115(3):211–252. https://doi.org/10.1007/s11263-015-0816-y.

Smelik, Ruben M., Klaas Jan De Kraker, Tim Tutenel, Rafael Bidarra, and Saskia A Groenewegen. 2009. A Survey of Procedural Methods for Terrain Modelling. In *Proceedings of the CASA Workshop on 3D Advanced Media in Gaming and Simulation (3AMIGAS)*, edited by Arjan Egges, Wolfgang Hürst, and Remco C. Veltkamp, pp. 25–34. CASA, Amsterdam, Netherlands. https://publications.graphics.tudelft.nl/papers/411, accessed February 3, 2026.

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press

Snitker, Grant, Jason D. Moser, Bobby Southerlin, and Christina Stewart. 2022. Detecting Historic Tar Kilns and Tar Production Sites Using High-Resolution, Aerial LiDAR-Derived Digital Elevation Models: Introducing the Tar Kiln Feature Detection Workflow (TKFD) Using Open-Access R and FIJI Software. *Journal of Archaeological Science: Reports* 41:103340. https://doi.org/10.1016/j.jasrep.2022.103340.

Sobotkova, Adela, Ross Deans Kristensen-McLachlan, Orla Mallon, and Shawn Adrian Ross. 2024. Validating Predictions of Burial Mounds with Field Data: The Promise and Reality of Machine Learning. *Journal of Documentation* 80(5):1167–1189. https://doi.org/10.1108/JD-05-2022-0096.

Somrak, Maja, Sašo Džeroski, and Žiga Kokalj. 2020. Learning to Classify Structures in ALS-Derived Visualizations of Ancient Maya Settlements with CNN. *Remote Sensing* 12(14):2215. https://doi.org/10.3390/rs12142215.

Soroush, Mehrnoush, Alireza Mehrtash, Emad Khazraee, and Jason A. Ur. 2020. Deep Learning in Archaeological Remote Sensing: Automated Qanat Detection in the Kurdistan Region of Iraq. *Remote Sensing* 12(3):500. https://doi.org/10.3390/rs12030500.

Southerlin, Bobby, B. Brilliant, Jason D. Moser, Grant Snitker, and Christina Stewart. 2021. The Archaeology of Tar and Pitch Production Sites in Francis Marion National Forest. *South Carolina Antiquities* 53:76–96.

Suh, Ji Won, Eli Anderson, William Ouimet, Katharine M. Johnson, and Chandi Witharana. 2021. Mapping Relict Charcoal Hearths in New England Using Deep Convolutional Neural Networks and LiDAR Data. *Remote Sensing* 13(22):4630. https://doi.org/10.3390/rs13224630.

Suh, Ji Won, and William Ouimet. 2023. Mapping Stone Walls in Northeastern USA Using Deep Learning and LiDAR Data. *GIScience & Remote Sensing* 60(1):2196117. https://doi.org/10.1080/15481603.2023.2196117.

Sung, Minsung, Jason Kim, Meungsuk Lee, Byeongjin Kim, Taesik Kim, Juhwan Kim, and Son-Cheol Yu. 2020. Realistic Sonar Image Simulation Using Deep Learning for Underwater Object Detection. *International Journal of Control, Automation and Systems* 18(3):523–534.

Togelius, Julian, Noor Shaker, and Mark Nelson. 2016. Introduction. In *Procedural Content Generation in Games*, by Julian Togelius, Noor Shaker, and Mark Nelson, pp. 1–15. Springer, Cham, Switzerland.

Touchstone, Samuel J. 1985. Charcoal-Tar Kiln. *North Louisiana Historical Association Journal* 16(2–3):103–105.

Trier, Øivind Due, Jarle Hamar Reksten, and Kristian Løseth. 2021. Automated Mapping of Cultural Heritage in Norway from Airborne Lidar Data Using Faster R-CNN. *International Journal of Applied Earth Observation and Geoinformation* 95:102241. https://doi.org/10.1016/j.jag.2020.102241.

United States War Department. 1944. *Field Fortifications*. US Government Printing Office, Washington.

US Geological Survey. 2021. *3D Elevation Program 1-Meter Resolution Digital Elevation Model (Published 20210129)—LA Sabine River LiDAR*. https://apps.nationalmap.gov/downloader/, accessed November 22, 2023.

Vadineanu, Serban, Tuna Kalayci, Daniël M. Pelt, and K. Joost Batenburg. 2024. Convolutional Neural Networks and Their Activations: An Exploratory Case Study on Mounded Settlements. *Journal of Computer Applications in Archaeology* 7(1):262–282. https://doi.org/10.5334/jcaa.163.

Verschoof-van der Vaart, Wouter B. 2022. Learning to Look at LiDAR: Combining CNN-Based Object Detection and GIS for Archaeological Prospection in Remotely-Sensed Data. PhD dissertation, Department of Archaeology, Universiteit Leiden, Leiden, Netherlands.

Verschoof-van der Vaart, Wouter, Alexander Bonhage, Anna Schneider, William Ouimet, and Thomas Raab. 2023. Automated Large-Scale Mapping and Analysis of Relict Charcoal Hearths in Connecticut (USA) Using a Deep Learning YOLOv4 Framework. *Archaeological Prospection* 30:251–266. https://doi.org/10.1002/arp.1889.

Verschoof-van der Vaart, Wouter B., and Karsten Lambers. 2019. Learning to Look at LiDAR: The Use of R-CNN in the Automated Detection of Archaeological Objects in LiDAR Data from the Netherlands. *Journal of Computer Applications in Archaeology* 2(1):31–40. https://doi.org/10.5334/jcaa.32.

Verschoof-van der Vaart, Wouter B., Karsten Lambers, Wojtek Kowalczyk, and Quentin P. J. Bourgeois. 2020. Combining Deep Learning and Location-Based Ranking for Large-Scale Archaeological Prospection of LiDAR Data from the Netherlands. *ISPRS International Journal of Geo-Information* 9(5):293. https://doi.org/10.3390/ijgi9050293.

Washburn, Albert L. 1988. *Mima Mounds: An Evaluation of Proposed Origins with Special Reference to the Puget Lowland*. Washington State Department of Natural Resources, Seattle.

Zakšek, Klemen, Kristof Oštir, and Žiga Kokalj. 2011. Sky-View Factor as a Relief Visualization Technique. *Remote Sensing* 3(2):398–415. https://doi.org/10.3390/rs3020398.

Cite this article: Peck, Katherine, Claudine Gravel-Miguel, Grant Snitker, and Matthew Helmer. 2026. Using Simulated Training Data to Locate Archaeological Sites with Machine Learning. *Advances in Archaeological Practice* 14(2):179–197. https://doi.org/10.1017/aap.2025.10130.

https://doi.org/10.1017/aap.2025.10130 Published online by Cambridge University Press