

Technical Note

A Technical Note on AI-Driven Archaeological Object Detection in Airborne LiDAR Derivative Data, with CNN as the Leading Technique

Reyhaneh Zeynali * , Emanuele Mandanici  and Gabriele Bitelli 

Department of Civil, Chemical, Environmental and Materials Engineering (DICAM), University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy; emanuele.mandanici@unibo.it (E.M.); gabriele.bitelli@unibo.it (G.B.)

* Correspondence: reyhaneh.zeynali2@unibo.it

Abstract

Archaeological research fundamentally relies on detecting features to uncover hidden historical information. Airborne (aerial) LiDAR technology has significantly advanced this field by providing high-resolution 3D terrain maps that enable the identification of ancient structures and landscapes with improved accuracy and efficiency. This technical note comprehensively reviews 45 recent studies to critically examine the integration of Machine Learning (ML) and Deep Learning (DL) techniques, particularly Convolutional Neural Networks (CNNs), with airborne LiDAR derivatives for automated archaeological feature detection. The review highlights the transformative potential of these approaches, revealing their capability to automate feature detection and classification, thus enhancing efficiency and accuracy in archaeological research. CNN-based methods, employed in 32 of the reviewed studies, consistently demonstrate high accuracy across diverse archaeological features. For example, ancient city walls were delineated with 94.12% precision using U-Net, Maya settlements with 95% accuracy using VGG-19, and with an IoU of around 80% using YOLOv8, and shipwrecks with a 92% F1-score using YOLOv3 aided by transfer learning. Furthermore, traditional ML techniques like random forest proved effective in tasks such as identifying burial mounds with 96% accuracy and ancient canals. Despite these significant advancements, the application of ML/DL in archaeology faces critical challenges, including the scarcity of large, labeled archaeological datasets, the prevalence of false positives due to morphological similarities with natural or modern features, and the lack of standardized evaluation metrics across studies. This note underscores the transformative potential of LiDAR and ML/DL integration and emphasizes the crucial need for continued interdisciplinary collaboration to address these limitations and advance the preservation of cultural heritage.

Keywords: archaeology; artificial intelligence; cultural heritage; feature detection; remote sensing



Academic Editor: Massimiliano Pepe

Received: 14 June 2025

Revised: 27 July 2025

Accepted: 5 August 2025

Published: 7 August 2025

Citation: Zeynali, R.; Mandanici, E.; Bitelli, G. A Technical Note on AI-Driven Archaeological Object Detection in Airborne LiDAR Derivative Data, with CNN as the Leading Technique. *Remote Sens.* **2025**, *17*, 2733. <https://doi.org/10.3390/rs17152733>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Archaeological feature detection plays a vital role in uncovering the secrets of the past, informing decision-makers to select more informed choices that balance development needs with the preservation of cultural heritage, advancing scientific knowledge, and engaging the public in exploring human history [1]. Across millennia, ancient settlements, roads, fortifications, and agricultural systems have left lasting marks on the landscape,

providing invaluable insights into past human behaviors, social structures, and technological advancements [2]. By carefully detecting and analyzing these features, archaeologists can piece together the complex network of ancient civilizations, shedding light on their history and cultural dynamics. Moreover, archaeological research provides crucial data that help scholars refine historical chronologies, understand trade networks, and reconstruct environmental changes that shaped human societies over time.

Yet, the preservation of archaeological sites faces countless threats, ranging from urban and agricultural development, looting, and wars to erosion and climate change [3–13]. Unregulated construction projects, mining activities, and expanding infrastructure often lead to the destruction of historically significant landscapes before their significance is even recognized [1,6,13–15]. The timely detection and documentation of archaeological features not only aid in safeguarding cultural heritage sites but also inform crucial land use planning and resource management decisions [2,14,16–18]. By identifying the locations of ancient settlements, burial sites, and other cultural features, potential damage from development projects can be mitigated, and conservation measures can be implemented to ensure the sustainable protection of heritage assets [2,13,19,20]. Moreover, archaeological feature detection contributes significantly to interdisciplinary research, bridging fields such as anthropology, history, geography, and environmental science [15,21,22]. The extensive dataset obtained from archaeological surveys and excavations helps test hypotheses, refine chronologies, and deepen our understanding of human–environment interactions over time. Furthermore, discoveries from archaeological feature detection enrich educational programs, museum exhibits, and public outreach initiatives, fostering curiosity and appreciation for cultural diversity among the general public [2,4,8,12,13,15]. By sharing the stories of past civilizations and the methods used to uncover their secrets, archaeologists inspire a sense of wonder and respect for our shared human history.

In their effort to uncover hidden features of ancient landscapes, archaeologists are increasingly turning to advanced geomatics technologies. Geographic Information Systems (GIS), Remote Sensing (RS) techniques such as LiDAR and Synthetic Aperture Radar (SAR), and artificial intelligence (AI) have become indispensable tools, enabling the analysis and interpretation of terrain features with unprecedented precision [4,10,23–27]. For instance, SAR technology offers unique capabilities for identifying buried archaeological features; examples include the case of Ostia–Portus in Italy, Uyük River Valley in Russia, and the Apamea site in Syria, where multi-band SAR successfully identified shallow-buried paleochannels, burial mounds, and looting pits, respectively [10,28,29]. To enhance the archaeological detection, SAR data can be fused with optical imagery. However, due to its complexity and need for further validation, its application is challenging [15,18,20,25,30]. These technologies facilitate large-scale archaeological surveys, reducing the need for invasive excavation methods and allowing researchers to explore sites that are otherwise inaccessible due to dense vegetation or challenging terrain. Among these technologies, LiDAR data and AI have emerged as potent tools for detecting archaeological features efficiently and accurately, offering invaluable insights into preserving endangered sites and discovering new archaeological areas [5,24,27,31,32]. Additionally, AI techniques such as Machine Learning (ML) and Deep Learning (DL) facilitate archaeological features' classification, identification, and segmentation, opening new avenues for archaeological research.

This technical note aims to critically examine recent archaeological studies that have employed derivatives of airborne (aerial) LiDAR point cloud data, such as Digital Elevation Models (DEMs) and ML techniques for archaeological site detection, highlighting the significance of these approaches and identifying critical research gaps and future research prospects. While various studies have explored the role of AI in archaeology and cultural heritage, they have primarily focused on different methodologies applied to various RS

datasets. For instance, [31] concentrates solely on employing DL approaches on diverse RS data (e.g., aerial photogrammetry, SAR, multispectral satellite imagery, and LiDAR) for digital preservation and object detection. Similarly, the focus of [1] is solely on one DL methodology (i.e., Convolutional Neural Network (CNN)), resulting in the review of six case studies using different RS data. The authors of [33] focused on state-of-the-art technologies related to AI and RS, while [34] primarily centers on Semantic Segmentation (SS) of point cloud data derived from LiDAR and photogrammetry for digital orthophotography mapping, damage investigations, object recognition, Building Information Model (BIM), and historical BIM. Furthermore, several reviews have concentrated on the application of LiDAR as well as other unmanned aircraft systems in archaeology and cultural heritage [35,36]. Moreover, [37] explicitly examines the airborne LiDAR data processing workflow. However, these reviews only partially focused on utilizing ML for archaeological feature detection from airborne LiDAR derivatives, which constitutes the primary focus of this study.

2. Research Aims

This study examines the current state of research on the integration of imagery derived from airborne LiDAR point cloud data and ML techniques in detecting archaeological features. Detecting such features is crucial in uncovering past civilizations and informing decisions that balance development with cultural heritage preservation. Airborne LiDAR technology, with its ability to penetrate vegetation and provide high-resolution topographic information, has transformed archaeological surveying by enabling the identification of subtle landscape features that may not be visible through traditional methods. Integrating ML algorithms with airborne LiDAR derivatives has the potential to enhance the efficiency and accuracy of detecting these features. This study aims to critically analyze the methodologies, applications, trends, and limitations reported in recent literature, with a particular focus on how ML has been employed to LiDAR-derived products in archaeological feature detection. It also seeks to identify existing research gaps and methodological challenges such as data processing, model generalizability, and the scarcity of labeled databases. By doing so, the paper aims to contribute to a more informed understanding of this interdisciplinary field and to suggest directions for future work.

3. Airborne LiDAR Technology in Archaeological Feature Detection

3.1. Introduction to Airborne LiDAR

The 21st century marks the initiation of using LiDAR data in archaeological exploration [38]. One of the earliest notable applications has been carried out in the UK, where LiDAR was utilized to identify and document earthwork traces of a Roman Fort in West Yorkshire, which traditional detection methods had overlooked [39]. Initially developed for terrain mapping and vegetation analysis, LiDAR's role has expanded to include the detection of hidden archaeological features, particularly in areas where traditional survey methods face limitations [38]. This growing recognition of LiDAR's capabilities has paved the way for its integration into archaeological research worldwide.

Airborne LiDAR technology is an active and non-invasive surveying method that utilizes laser scanning to generate highly detailed three-dimensional (3D) maps of the terrain surface, creating detailed 3D point clouds over vast areas. A standard airborne LiDAR setup comprises Airborne Laser Scanning (ALS), aircraft positioning using a Global Navigation Satellite System (GNSS), and an Inertial Measurement Unit (IMU) [23]. A laser scanner, typically mounted on an aircraft like a plane, helicopter, or drone, emits pulses, generally in the Near-Infrared Range (NIR), at average frequencies of around one million pulses per second (1 MHz) in various directions along the flight path toward the ground [40].

The initial commercial systems operated at 10 kHz and were bulky in size, while the contemporary systems are smaller, lighter, and can handle multiple laser returns [41]. For each laser pulse that hits the surface, discrete return LiDAR systems [24,40] detect and record a limited number of returns, while full-waveform systems [35,38] record the entire backscattered energy profile (continuous signal). These pulses bounce off objects such as the ground surface, vegetation, and buildings, with their positions determined by calculating the time delay between emission and reception of each echo, along with the direction of the laser beam and the scanner's position [5,32]. In addition, bathymetric LiDAR has emerged as a specialized tool for underwater surveying. Unlike NIR LiDAR, which utilizes larger wavelengths, bathymetric LiDAR employs smaller wavelengths in the green spectrum to penetrate the water column effectively [42]. While early iterations of bathymetric LiDAR faced limitations in return point density and spatial resolution, recent advancements have significantly improved instrument quality [43]. This progress has enabled a wide range of archaeological applications, including documenting submerged sites.

3.2. Advantages of Airborne LiDAR in Archaeology

Airborne LiDAR's primary application in archaeology has been identifying archaeological structures visible as topographic imprints on the ground surface [39]. This technology has revolutionized archaeological surveying, enabling researchers to identify and analyze features such as ancient structures, roads, and burial mounds with improved accuracy. Airborne LiDAR offers a significant increase in spatial resolution compared to photogrammetric and satellite-derived products obtained from stereo or tristereo imagery [44]. On the other hand, Unmanned Aerial Vehicle (UAV) photogrammetric surveys can be severely limited by the presence of vegetation and by georeferencing challenges, especially when establishing ground control points in hard-to-reach areas [45]. LiDAR allows for the filtering of vegetation, buildings, and other man-made structures to create high-resolution bare-soil Digital Terrain Models (DTMs), which can aid in the detection of archaeological remains. However, in areas with sparse ground returns, such as dense vegetation or urban settings, the ground surface must be interpolated, which can reduce the quality of DTMs [16,37] and, as a result, the accuracy of archaeological feature detection. The combination of LiDAR-based DTMs and different visualization techniques has significant contributions in both previously overlooked, due to dense vegetation cover, and well-studied archaeological areas (Figure 1) [23]. Besides its efficiency in data acquisition, airborne LiDAR can support the identification of buried or partially buried archaeological features such as ancient trenches, roadways, and agricultural fields [28–31], as well as geoarchaeological features in low-relief alluvial landscapes [46]. Additionally, for some countries, airborne LiDAR data are already accessible online immediately or upon request, offering economic benefits as they are more cost-effective and efficient than traditional methods [47].

Airborne LiDAR technology can reveal not only visible archaeological structures but also hidden features that are often difficult to detect using traditional methods. These hidden features, such as buried structures, roads, and agricultural fields, are of great significance in archaeological research [48]. By revealing well-preserved sites obscured by natural factors such as dense vegetation or soil erosion, LiDAR enables archaeologists to explore previously overlooked regions and gain a deeper understanding of human history and culture [49]. This aspect of LiDAR technology highlights its potential to revolutionize the study of civilizations that have long since disappeared, offering new interpretations of their social, economic, and environmental systems [48,50]. Some of the most significant findings facilitated by airborne LiDAR technology in archaeology include the mapping of hidden cities and vast urban complexes, such as the expansive Maya civilization in Central America [24] and the elaborate network of Angkor in Cambodia [51,52], as well as evidence

of the architectural sophistication of ancient civilizations, showcasing elaborate structures like earthworks [6], complex road networks [53] used for trade and communication, hidden agricultural fields [2], and defensive structures [31]. Airborne LiDAR has significantly enhanced the visualization and mapping of ancient cities and extensive landscapes previously obscured by dense vegetation or sediment accumulation.

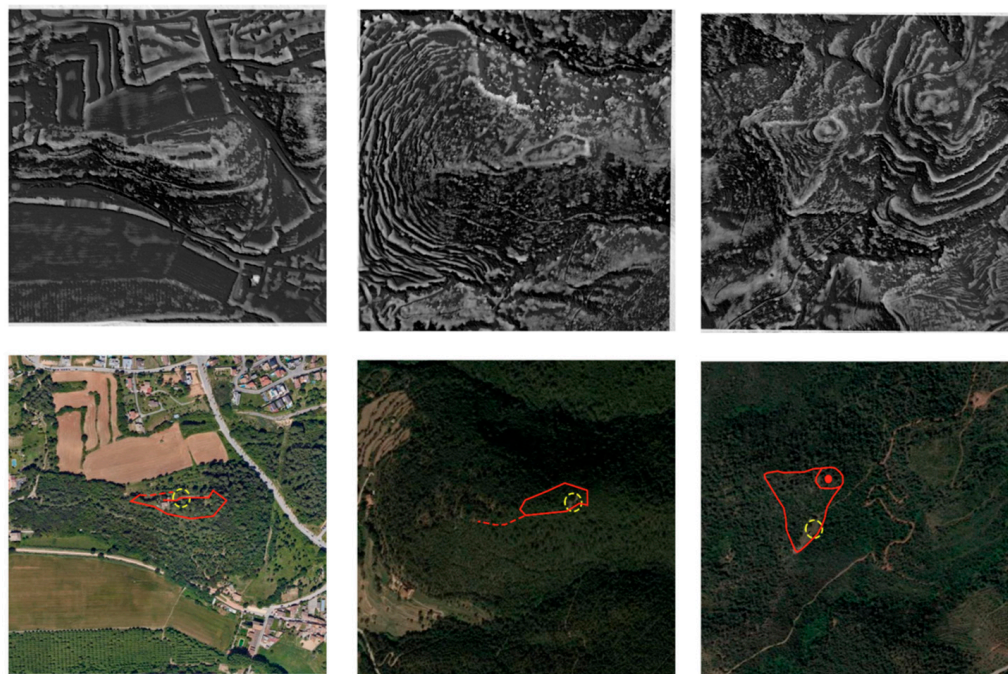


Figure 1. Multi-scale relief visualization (**top**) of digital terrain model and satellite view (**down**) of three archaeological sites: Puig d'en Rovira (**the left column**), Puig Castell (**the middle column**), and Torre Roja (**the right column**), Spain. In the satellite images, the walls are marked in red and the known locations of the settlements in yellow [23].

3.3. Challenges and Limitations of Airborne LiDAR in Archaeology

Airborne LiDAR has transformed archaeological prospects; however, its implementation faces challenges. The high cost of data acquisition and limited global availability restrict accessibility, particularly in less industrialized regions [36,40]. Collecting airborne LiDAR can cost two orders of magnitude more than acquiring equivalent commercial multispectral satellite data [25]. UAV LiDAR can reduce costs but is limited in coverage area [12,25]. Furthermore, LiDAR surveys can be challenging due to logistical and regulatory constraints [48]. Drone usage, for instance, requires navigating varying national legislations and obtaining special permits for restricted areas like airport surroundings [37,48]. Weather, vegetation seasonality, and dense ground cover can affect data quality, while resolution and point density influence the detectability of small features. Weather conditions like heavy rain, fog, dust, or haze can interfere with laser beam measurements, leading to inaccuracies [31,48]. The seasonal variability of vegetation, particularly the leaf-on and leaf-off conditions, significantly impacts the effectiveness of LiDAR in penetrating foliage and revealing subtle archaeological features [29,36,48]. This necessitates careful planning of acquisition dates, which is often not possible when relying on openly available datasets acquired for different purposes [36]. Exceptionally dense or multi-layered forests and tall grasses can still obscure archaeological features, making it difficult to separate ground returns from vegetation returns [12,25,36,48]. This can result in rough surfaces in DEMs that mask underlying features. Another issue is the need for high-quality LiDAR, which requires significant computational resources and expertise, and such data are not always

accessible [16,27]. The quality of LiDAR data is crucial for accurate and reliable results in various applications, with point density as a key parameter to assess quality. Point density, representing the number of LiDAR points per unit area, can directly influence the level of detail and precision of derived products [47,51]. Ensuring adequate point density and transparency in reporting are essential for maximizing the utility and trustworthiness of LiDAR-derived information across diverse applications [54]. While advancements allow for higher densities, lower resolution datasets (e.g., one point per square meter) can lead to artifacts in the data or omit smaller features [16,27,55]. Regular updates and monitoring of surveyed and mapped areas are also necessary [40]. Other challenges and limitations to this technology include high initial investment costs, managing and interpreting large data volumes, vulnerability to weather and environmental conditions affecting data collection, as well as logistical obstacles related to site access, permissions, and data privacy [48]. It furthermore requires collaboration between archaeologists and RS experts, ensuring a more comprehensive and deeper data analysis.

4. Machine Learning in Archaeological Feature Detection

4.1. Overview

Methods for detecting archaeological features have evolved from subjective approaches, which rely primarily on personal interpretation without clear and standardized criteria [56], to more scientific ones, aided by computing technology. Archaeological feature detection involves field surveys, RS, and integrating RS with AI [23,33]. The fusion of RS data with AI marks a new era in archaeology. ML algorithms, a subset of AI, have become indispensable tools for processing and analyzing large volumes of data in archaeological contexts. Traditional techniques often suffer from limited coverage, subjective interpretation, and unquantified error rates, leading to false positives (FPs) (detecting features that are not actually archaeological) and false negatives (FNs) (missing actual archaeological features) [4,7,14,50], which ML blended with RS can overcome by creating a systematic solution [33]. ML is broadly defined as the capacity of intelligent systems to learn and improve from prior data [1]. It involves optimizing a model to transform input into a desired output with increasing efficiency [57,58]. This optimization process, known as training, is performed using a relevant set of training data. ML algorithms are trained to derive mathematical classifiers or feature vectors and apply them to extract, sort, classify, and interpret new data [57]. AI uses such ML models to automate decision-making and data analysis. ML has been successfully applied in various domains of computer science, including computer vision, data classification, knowledge extraction, and speech recognition [1,57,58]. They can be applied to a range of digital data, with the most common types being numerical/categorical data, images, and geospatial data [4,17,57].

ML methods are categorized into supervised learning, unsupervised learning, and reinforcement learning. A fourth category, semi-supervised, is also sometimes mentioned [1,58]. The fundamental difference often lies in the type of training data and whether these data have known outputs or labels [58,59]. Supervised learning involves training models using data with known labels to learn a mapping function that can accurately predict the output for new, unseen inputs [1,58]. This process involves iteratively refining the model based on the provided training data. Common supervised learning tasks are divided into two main groups: regression and classification. Regression tasks aim to learn a real-valued function for continuous outputs, while a classification task assigns input elements to a predefined set of discrete categories or labels [58]. Examples of common supervised algorithms include decision trees, maximum likelihood, Random Forests (RF), Support Vector Machine (SVM), Artificial Neural Networks (ANNs), and DL models like CNNs [7,60–63]. SVM, maximum likelihood, and ANNs are commonly used for classifying raster and images with training

data that are created by drawing polygons around known objects [60,62]. Conversely, in unsupervised learning, the objective is to find structure, patterns, and inherent groupings within the unlabeled data autonomously [1,58]. Unlabeled data help the model learn and gain familiarity by attempting to make predictions without knowing the true targets [59], and they can be optimal when features of interest are not previously known [53]. Clustering algorithms are among the most widely used methods in archaeology and cultural heritage for discovering patterns and structures in unlabeled datasets [64]. The simplest and most commonly used clustering algorithm is K-means [64], which is also integrated into advanced architectures like PointNet++ for hierarchical point grouping [65]. An architecture refers to the specific design and structure of a neural network. It encompasses the arrangement of layers, the types of layers (e.g., convolutional, pooling, fully connected), the connections between layers, and the methods used for processing data. Different architectures are tailored to solve various tasks such as image classification, object detection, or segmentation, and they influence the model's performance and efficiency. Reinforcement learning, with its common Q-learning algorithm, trains models through rewards and penalties [1,66]. Finally, semi-supervised learning aims to enhance learning performance by utilizing both labeled and unlabeled data. Its common implementation strategy involves a two-stage process. First, an unsupervised pretraining model is performed using a large quantity of unlabeled data, followed by a supervised fine-tuning phase using a more limited set of labeled data [18,58]. The primary motivation for employing semi-supervised ML models is to overcome the challenge of limited labeled training data, a common issue in many domains, including cultural heritage and archaeology.

4.2. Application of Machine Learning in Archaeological Feature Detection

ML techniques have become valuable tools for archaeological feature detection, providing efficient alternatives to manual analysis of large RS datasets. Common ML algorithms applied in this field include RF and SVM [7,11,25,62,67]. RF has been effectively used for tasks such as classifying landscape features to identify Viking ring fortresses [11] and mapping ancient canals [25], valued for its robustness, efficiency, and ability to handle various data distributions [11,25]. SVM algorithms have demonstrated utility in detecting historical terrain anomalies and classifying stone-walled structures [12,60,62]. Beyond these, semi-automated methods like template matching and Object-Based Image Analysis (OBIA) are also employed, particularly successful for features with simple geometric shapes like circles and rectangles, which are frequently found in archaeological records but are rare in natural formations [4,7,12,54,68]. While ML significantly boosts efficiency and can achieve high true-positive rates in specific contexts, challenges remain, including sensitivity to labeling errors and limitations in generalizing to diverse landscapes or complex, heterogeneous archaeological structures altered over millennia [7,54]. The performance of these methods can also be significantly influenced by terrain type and the inherent quality of the LiDAR data [69].

4.3. Deep Learning

DL is an advanced form of ML originated in the 1940s with the goal of mimicking human brain functions [70]. After facing challenges like overfitting and limited data, DL regained popularity in 2006 due to its significant advancements achieved in speech recognition [31]. The term *deep* refers to the development of NNs with more than one hidden layer [1,59,63,70]. Deep Neural Networks (DNNs) facilitate automated learning by iterating through massive amounts of data in images, text, or videos [71]. They seek to exploit the unknown structure in the input data distribution to discover good representations, often at multiple levels [70]. In other words, DL models build understanding in a hierarchical way,

where complex features are composed of simpler ones learned in earlier layers and without humans needing to define those features explicitly [1,63,70]. DNNs are composed of simple, highly interrelated processing units called neurons, organized in multiple layers [70]. There is an input layer, one or more intermediate or hidden layers, and a final output layer [63]. The connections between neurons are represented mathematically by trainable parameters called weights [70,72]. The training process begins with the input of raw data, such as images, text, video, or spatial datasets like DTMs, Local Relief Models (LRMs), or 3D point clouds [20,30,38,54–57]. During forward propagation, these data are passed through multiple layers of the network, where each neuron computes a nonlinear activation on the weighted sum of its inputs and transmits the signal forward [14,17,24,63,70,73]. The most commonly used activation function is Rectified Linear Units (ReLU). ReLU is an activation function used in neural networks, defined as $f(x) = \max(0, x)$. It introduces non-linearity into the model, enabling it to learn complex patterns by allowing only positive values to pass through while setting all negative values to zero. This helps in addressing issues like vanishing gradients and improving training efficiency [17,24,73]. As the data progresses through the hidden layers, the network performs feature extraction and automatically learns to detect increasingly complex patterns [1,59,63]. This is a key distinction of DL from classical ML. The final output layer generates predictions depending on the task: this may include class probabilities [3,19], classifications [3,19], or segmentations [9]. The goal of training is to adjust the model's weights to minimize the difference between the model's output and the desired output for a given task [3,8,70]. After each forward pass, a loss function measures prediction error [70,73]. Then, back propagation propagates the loss information backward through the network and computes gradients, which indicate the contribution of each weight to the error [24,70]. It is important to note that not all DL rely on backpropagation due to limitations in scalability or hardware constraints [15,33,53,66]. An optimizer then uses these gradients to adjust the trainable weights, aiming to minimize the loss function [8,70,73]. Training runs in steps called epochs, where each one involves a full forward and backward pass of all training data through the network [2,70]. After each epoch or a set of epochs, model performance is validated using unseen data to ensure generalization and prevent overfitting [2,59].

DL is a step further within ML, employing models such as ANNs, CNNs, and RNNs, which follow common methodologies like classification and regression and can significantly improve accuracy with increased data availability [1,3,63,74]. However, it requires more extended training but shorter inference times than traditional ML algorithms, relying on large datasets for better accuracy [57]. A key difference between classical ML techniques and DL is that while the former often need feature engineering (i.e., selecting, transforming, or creating new features from raw data), human experts to carefully select input features (such as spectral indices), and the prior calculation and determination of a range of possible statistically significant input features [6], the latter performs feature extraction by itself [1]. In other words, DL algorithms can identify and extract meaningful patterns or features directly from raw data, which can be particularly time-saving and advantageous when working with complex and high-dimensional datasets [1,63]. Similar to ML, DL encompasses different types of learning (i.e., supervised, unsupervised, reinforcement learning, self-supervised) [1,58]. However, DL models generally require considerable computational resources [6] and large amounts of data to achieve higher accuracy [3,74]. Traditional ML methods might be more applicable or sufficient when dealing with limited datasets or specific cases where targets have limited spectral and geometric variations [63].

Based on the task, DL models can be categorized by their output. First are object detection models that identify and locate specific objects within an image, typically drawing a bounding box around them and providing a probability of the object's presence [3,68].

Examples of object detection architectures include Region-based CNNs (R-CNNs), Faster R-CNN, YOLOv8, YOLOv8-CDD, YOLOv5, and YOLOv4 [3,18,68,75–77]. Secondly, SS models classify each pixel of an image with a corresponding class, defining semantic regions and segmenting them [24,68]. This provides both the classification and position of archaeological structures [3]. Examples of SS architectures include Faster R-CNN, U-shaped CNN (U-Net), Mask R-CNN, ResUnet, and Fully Convolutional Networks (FCNs) [3,68]. CNNs are among the most studied architectures, especially when the input data are images and videos [1,3,14,74]. They have been successfully used in various image applications, including image classification, object recognition, video classification, and scene labeling [3,74]. Different CNN architectures exist with variation in their layer organization [24]. Figure 2 demonstrates an example of a simple CNN architecture, in which the input image passes through multiple convolutional layers that extract local features such as edges and textures [3,74]. The pooling layers help reduce spatial dimensions and control overfitting by preserving the most relevant features. The extracted features are then fed into fully connected layers that interpret the information for classification. Finally, an activation function called SoftMax is applied to the classes. Softmax is an activation function used in neural networks, particularly in the output layer for classification tasks. It converts raw scores (logits) into probabilities by exponentiating each score and then normalizing them so that their sum equals one. This allows the model to assign a probability to each class, facilitating multi-class classification. This assigns probabilities to different classes, enabling the network to learn hierarchical representations of data automatically [59,74]. A more complex CNN architecture can be found in Figure 3. It illustrates the architecture, developed to segment clearance cairns from LiDAR datasets. U-net enables pixel-level classification by combining a contracting path for feature extraction with an expanding path for precise localization. The contracting path performs a series of convolution and max-pooling operations to capture contextual features, while the expanding path restores spatial resolution through up-sampling and concatenates high-resolution features from earlier layers. This architecture is especially effective with limited training data and allows for the detailed extraction of archaeological features through a sigmoid activation function in the final prediction layer [2].

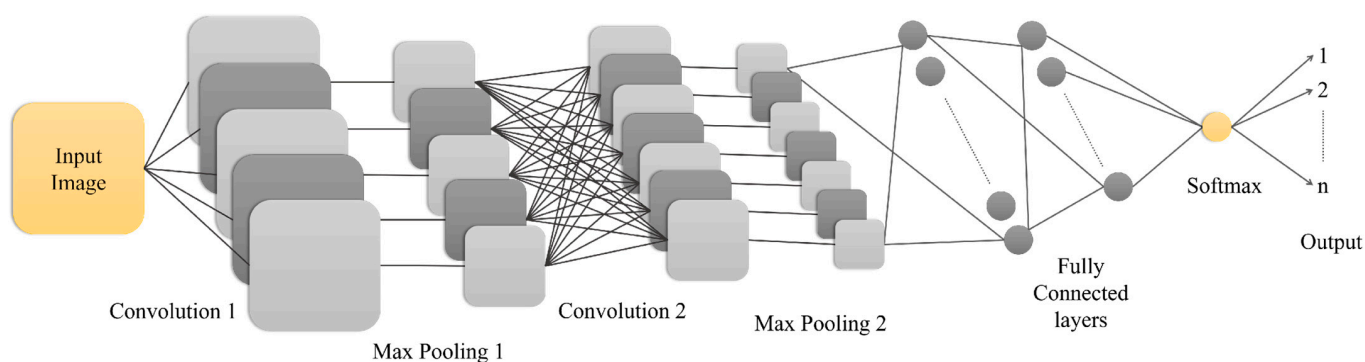


Figure 2. Structure of a simple Convolutional Neural Network (CNN) architecture for n-class classification. The input image undergoes two convolutional and pooling operations, extracting hierarchical features. The resulting feature map is flattened and passed through two fully connected layers. Finally, a Softmax activation function is applied to the output of the last fully connected layer, converting it into a vector of n probabilities corresponding to the classification classes.

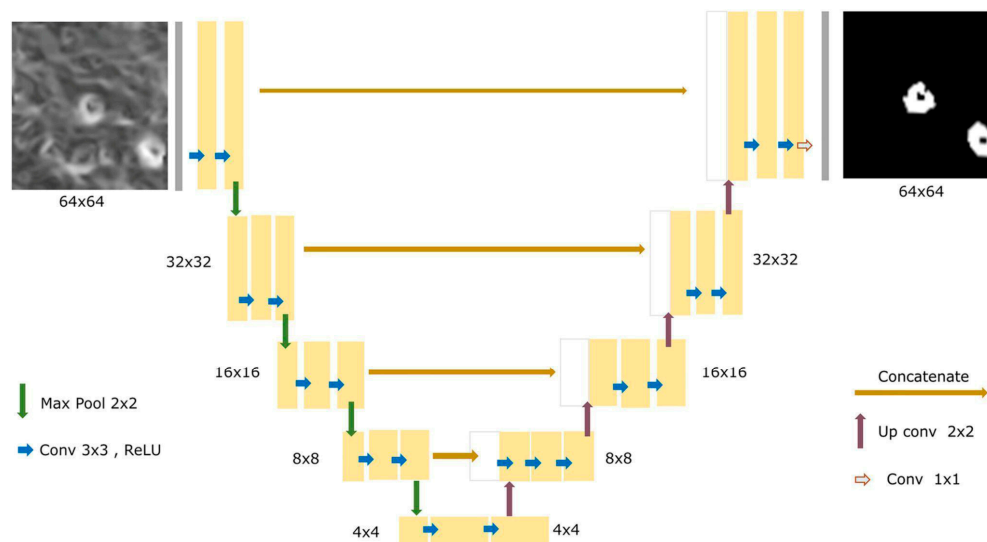


Figure 3. U-shaped Convolutional Neural Network (U-net) architecture used for the segmentation of clearance cairns from LiDAR data [2]. The network consists of a contracting path (down-sampling) that extracts contextual features through convolution and max-pooling, and an expanding path (up-sampling) that enables precise localization via up-convolution and concatenation with corresponding feature maps. This structure allows for pixel-level classification and accurate delineation of archaeological features, even with limited training data.

4.4. Application of Deep Learning in Archaeological Feature Detection

DL has emerged as a transformative and increasingly adopted approach for automated archaeological feature detection, particularly leveraging LiDAR data due to its capacity to process immense datasets and learn intricate patterns. These methods predominantly utilize CNN architectures, including R-CNNs for both localizing and classifying multiple objects [15,20,68,78]. Specialized variants like Faster R-CNN have been applied to detect barrows, Celtic fields, charcoal kilns, and pitfall traps [10,13,15,24], while Mask R-CNN is used for detailed pixel-level segmentation of features such as charcoal hearths and shell rings [10,22,24,32]. SS models like U-Net and DeepLabv3+ are widely used for identifying ancient construction activity, agricultural terraces, and historical land use features like relict charcoal hearths and stone walls at a pixel level [5,24,61,68,79,80]. Furthermore, YOLO (You Only Look Once) frameworks (YOLOv3, YOLOv4, YOLOv5, YOLOv8, and YOLOv8-CDD) are favored for their speed in object detection, with applications ranging from shipwreck detection to identifying burial mounds and other archaeological features [18,42,75–78]. DL models consistently achieve high accuracy and detection rates [2,7]. TL and data augmentation are crucial strategies that address the challenge of limited labeled archaeological training data and enhance model generalization capabilities [14,26,42]. Despite these advancements, DL still faces hurdles, including demanding computational resources and a persistent high rate of FP due to the morphological similarities between archaeological features and natural or modern landscape elements [15,29,69,78,80]. To mitigate FP, techniques such as Location-Based Ranking (LBR) and integrating background topography into model training have been developed [15,42,51,75]. Fundamentally, these automated detection methods are designed not to replace human experts, but to serve as an efficient instrument in the archaeologist’s toolkit, enabling rapid mapping over extensive areas and allowing experts to reallocate time to in-depth analysis and field validation.

4.5. Transfer Learning

Although CNNs typically perform best with large amounts of labeled data, in many real-world scenarios, such data may be limited. In these cases, transfer learning (TL) can

offer an effective approach to mitigate this challenge. TL is a technique that applies the knowledge gained from solving one problem to a different but related problem [3]. The core idea is to reuse a model that has been pre-trained on a large amount of data (for instance, on the ImageNet dataset) for a source problem to solve a target problem where only limited data are available [3]. However, differences between the characteristics of the original training data (for instance, standard computer vision images) and the new data (such as aerial LiDAR-derived images with different scale and rotation properties) need to be considered [14]. TL is typically achieved by retraining only a few selected layers of the pre-trained model on the new dataset [3,14]. Common strategies for applying TL include removing the final layer of the pre-trained network and replacing it with a new layer suited for the target classification categories, or fine-tuning the model by adding new training parameters [14]. In fine-tuning, either all model weights can be updated, or the weights in the lower layers can be frozen while only updating the upper layers [14]. Studies indicate that TL can significantly improve model accuracy in situations where the available data for training is limited [2,3,10,74]. This technique directly addresses a major challenge in DL, which is the requirement for vast amounts of labeled training data to achieve higher accuracy [1,3,74]. While commonly associated with CNNs, TL can be applied to various ML paradigms, including supervised and unsupervised learning [42]. Its advantages include reduced training time and data requirements, improved generalization, and the ability to address domain shift [72]. However, its effectiveness depends on the similarity between the source and target domains, and there is a risk of overfitting, especially when the target dataset is small or significantly different from the source dataset [14,72,81]. Despite these limitations, TL remains a valuable technique for accelerating model development, improving performance, and addressing challenges in diverse ML applications, particularly in archaeological contexts with data limitations and small sample sizes [14,42,72,81].

5. Past Research Applying Machine Learning on Airborne LiDAR Derivatives for Archaeological Feature Detection

Over the past four decades, advancements in RS, ML, and cloud computing have revolutionized our ability to explore ancient landscapes, with airborne LiDAR technology playing a pivotal role in improving spatial resolution for surveys of previously inaccessible forested areas [25]. The systematic approach for applying ML techniques to airborne LiDAR data for archaeological feature detection starts with carefully collecting and pre-processing LiDAR and archaeological datasets, ensuring data quality and compatibility. Feature engineering follows, wherein relevant attributes are extracted and engineered to enhance the model's ability to identify archaeological features from the LiDAR derivatives. A critical decision point arises in model selection, where the choice between traditional ML algorithms and DL architectures is made, with the possibility of including TL on pre-trained models. The subsequent steps involve rigorous training and evaluation of the selected model, ensuring its robustness and generalization to unseen data. Once validated, the model is deployed for archaeological feature detection on new airborne LiDAR datasets. Continuous monitoring of the model's performance in real-world applications enables improvements as necessary, thereby ensuring the efficacy of the archaeological research efforts (Figure 4).

The core idea of each ML/DL model is the ability to perform well on data that has not been seen during training [4,59]. Ensuring and evaluating this ability, known as generalization, involves several techniques. The foundation of good generalization lies in having a sufficient amount of high-quality, representative training data [24,26,47]. However, collecting large, labeled datasets can be challenging, particularly in specialized domains like archaeology [1,12,14,29,41,57]. Furthermore, using a separate validation dataset [40]

as well as appropriate evaluation strategies [82] could ensure generalization. Standard evaluation measures developed for general object detection tasks may not be suitable for archaeological applications due to the unique characteristics of archaeological objects and the importance of geospatial information for archaeologists [82]. Another strategy to increase the robustness of the model is applying data augmentation [3,10,40,78]. This involves creating artificial training samples by applying various transformations (such as cropping, flipping, rotation, scaling, and shifting) to the original data [78]. Applying multi-directional hillshade or other visualization techniques to DEMs can also augment the training data [3,26]. For 3D point clouds, combining different augmentation methods, such as Gaussian noise and random rotation, has shown potential for improving DL models with small datasets [3]. TL can also help with generalization [14,57,72]. For instance, [14] successfully reapplied the TL of a deep CNN, initially trained on Lunar LiDAR datasets, for detecting historic mining pits [14]. Choosing an architecture that can handle the multi-scale nature of archaeological features or integrate multiple data sources [7,29,40,54] and optimizing the model's hyperparameters (such as the number of epochs) also enhances the model's ability to generalize to unseen data [2,82,83]. Hyperparameter selection in DL is challenging due to the lack of universal guidelines and the high context dependency on data type, task, and model architecture [2,14]. This process is also computationally expensive and time-consuming, often requiring trade-offs that limit full optimization [7,17,63,81]. Poor choices can lead to overfitting [6,14,18,20,54], underfitting [54], convergence issues [15,22,61], and biased results [20,53], especially in cases of small or unbalanced datasets. Evaluating tuning outcomes is further complicated by issues like data imbalance and multi-objective goals [20,70]. To overcome these challenges, systematic tuning methods such as cross-validation [2,14,75], random search [14,17,54], early stopping [54,68], and automated learning rate selection [2,54] can be applied. Data-centric strategies like TL [2,4,20,57,58,72,81], data augmentation [8,18,20,54,72,78], normalization and pre-processing [22,31,58,84], dataset balancing [13,20,53,54], and high-quality and diverse training data [17,24,42,57] can help improve generalization and robustness. Choosing simpler or specialized architectures suited for small datasets [3,5,7,14,61,63], and adapting models for single-channel inputs like LiDAR data [14], enhances performance. Finally, integrating domain experts in the process ensures higher data quality and model reliability [5,16,17,24,26,57], making hyperparameter tuning a manageable but essential step in successful DL applications.

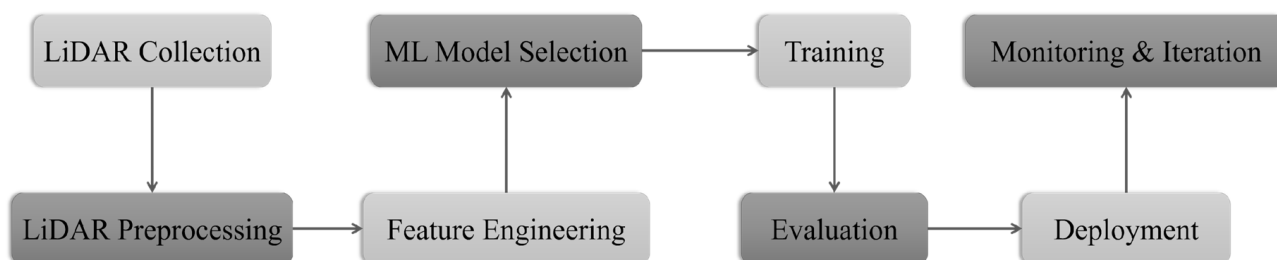


Figure 4. Steps involved in applying Machine Learning (ML) to airborne LiDAR data for archaeological feature detection (the workflow illustrates a typical pipeline highlighting the interdisciplinary nature of the process).

Integrating airborne LiDAR derivatives with ML algorithms for archaeological feature detection has been the subject of 45 case studies (locations of the studied area are demonstrated in Figure 5), found after internet research. Figure 6 displays a histogram representing the yearly distribution of these articles. This trend suggests a growing interest in the research topic, particularly around 2021 and 2023. On the Scopus website, a set of

keywords was searched within the title, abstract, and keywords of review and scientific articles using the following advanced query:



Figure 5. Regional research biases of the case studies performed to detect archaeological features using airborne LiDAR data and machine learning methods. Each red circle represents a case study conducted in that geographic area.

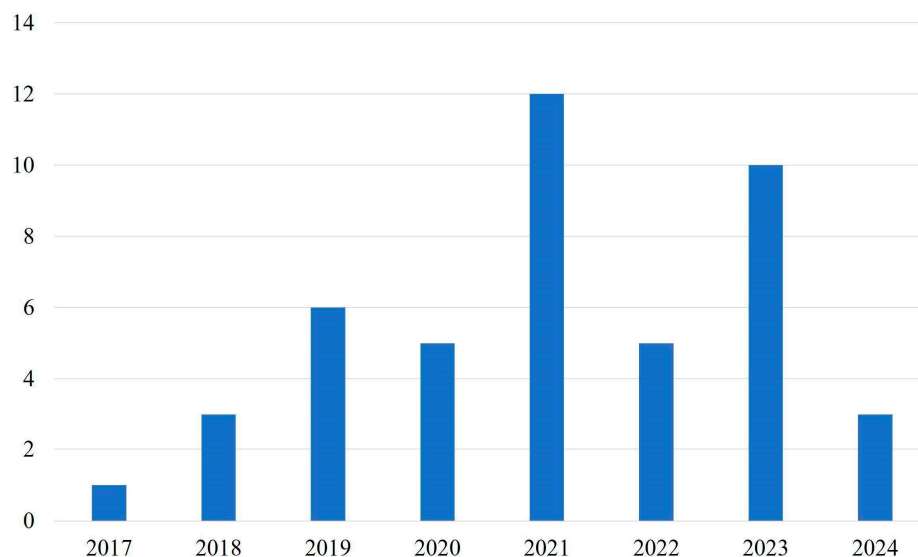


Figure 6. Accelerating research focus on archaeological feature detection using airborne LiDAR and machine learning for the analyzed articles.

TITLE-ABS-KEY (“deep learning” OR “machine learning” OR “artificial intelligence” OR “AI” OR “semantic segmentation” OR “computer vision”)
 AND TITLE-ABS-KEY (“archaeolog*” OR “historical “ OR “cultural” OR “heritage”)

AND TITLE-ABS-KEY (“airborne” OR “aerial”) AND (“LiDAR” OR “laser” OR “point cloud”))

AND TITLE-ABS-KEY (“object” OR “site” OR “pattern” OR “feature”) AND (“detection” OR “extraction” OR “recognition”))

AND (LIMIT-TO (DOCTYPE, “ar”) OR LIMIT-TO (DOCTYPE, “re”))

AND (LIMIT-TO (LANGUAGE, “English”)).

This search resulted in 41 articles. After a detailed screening, 15 were excluded for reasons such as duplicate records, irrelevant use of AI or ML, and absence of feature detection/extraction methods or insufficient methodological detail relevant to the scope of this work. The remaining 26 articles are presented in Table 1. The same keywords were used to search Google Scholar within the same timeline (between 2017 and 2024) as the previous search. This search yielded 34 articles, 15 overlapping those on the Scopus website. The remaining 19 new articles were also added to Table 1, resulting in a total of 45 articles. Of these, 32 used deep CNNs for automatic object detection, applying different architectures such as U-Net, YOLO, CarcassonNet, WODAN, VGG, Deeplab, and DeepMoon. However, as highlighted in some papers, their application in large-scale archaeological mapping may require further evaluation and extension, including the development of optimal training sample selection methods and evaluation of LiDAR-derived data as inputs [38,72,82]. One article applied a multimodal segmentation DL model known as CMX (i.e., an RGB-X SS with Transformers), a fusion of SS with Transformers [29]. Another study conducted a comparison between ML (RF) and DL (fully connected networks) [54]. The remaining 11 case studies apply different ML algorithms such as SVM, unsupervised ISODATA, RF, and template matching classifiers to detect the archeological features in airborne LiDAR data (Table 1). To improve readability, Table 1 has been subdivided into three sections, Table 1 (a), Table 1 (b), and Table 1 (c), based on the spatial resolution of the LiDAR-derived products used in the respective ML/DL methods.

Table 1. Case studies that applied different artificial intelligence methods on airborne LiDAR derivatives to detect archaeological features. (a) Derivatives with a resolution lower than 0.5 m. (b) erivatives with a resolution higher than 2 m. (c) erivatives with unexpressed metric resolution.

Authors	Archaeological Sites/Objects	Study’s Location (Extent)	LiDAR Derivative and Resolution	Detection Method (Architecture/ Algorithm)	Quality Evaluation
(a)					
[8]	Ancient City Walls	Jinancheng, China (16 km ²)	0.5 m DEM	CNN (U-Net segmentation)	Precision 94.12%
[61]	Ancient Agricultural Water Harvesting Systems (Terrace and Sidewall)	Central Negev Desert, Israel (1800 km ²)	0.125 m DTM; 2 points/m ²	CNN (modified U-Net)	IoU 53%
[32]	Pitfall Systems	Suomenselka, Finland (6778.9 km ²)	0.25 m DEM; 5 points/m ²	CNN (-)	Reliability 80%
[5]	Tar Production Kilns	Kuivaniemi (2760 km ²), Hossa (2004 km ²), and Näljänkä (2304 km ²), Finland	0.25 m DEM; 5 points/m ²	CNN (U-Net)	Accuracy 93–95% Precision 82–97% Recall 72–99% F1-score 77–97%

Table 1. Cont.

Authors	Archaeological Sites/Objects	Study's Location (Extent)	LiDAR Derivative and Resolution	Detection Method (Architecture/ Algorithm)	Quality Evaluation
[60]	Precolonial Stone-Walled Structures (Circular Homestead, Agricultural Terrace, and Road)	Thaba-Chweu, South Africa (31.25 km ²)	0.2 m DTM	ML (Support Vector Machine)	Accuracy 95%
[25]	Ancient Canals (Maya Wetland)	Rio Bravo, Belize (~5 km ²)	0.5 m DEM	ML (Random Forest)	Accuracy 66%
[19]	Linear Structures (Embankment, Ditch, Hollow Path, etc.)	Blois, France (270 km ²)	0.5 m DTM	ML (Support Vector Machine)	-
[82]	Barrows and Celtic Fields	Gelderland, The Netherlands (2200 km ²)	0.5 m DTM; 6–10 points/m ²	Faster Region-based CNN	-
[17]	Historic Stone Walls	Aro, Denmark (88 km ²)	0.4 m DTM	CNN (U-Net segmentation)	Accuracy 93%
[41]	Celtic Fields and Burial Mounds	The Białowieża Forest, Poland (697.8 km ²)	0.5 m DTM; 11 points/m ²	CNN (U-Net)	F1-score 58% IoU 50%
[72]	Topographic Anomalies	Brittany, France (200 km ²)	0.5 m DTM; 14 points/m ²	TL Mask Region-based CNN (ResNet-101)	Accuracy <77%
[13]	Grave mound, Pitfall trap, Charcoal Kiln	Norway (937 km ²)	0.5 m DTM; 5 points/m ²	Faster Regional based-CNN	Accuracy ~70%
[53]	Trace Hollow Roads	Veluwe, The Netherlands (93.75 km ²)	0.5 m DTM	CNN (CarcassonNet)	Accuracy 89%, F1-score 42%
[22]	Relict Charcoal Hearth Sites	Germany (3.4 km ²)	0.5 m DEM	Modified Mask Region-based CNN	Recall 83%, Precision 87%
[80]	Ancient Agricultural Terraces and Walls	Negev, Israel (-)	0.1 m DTM	CNN (U-Net segmentation)	Precision (Terrace 87%, Wall 60%)
[15]	Barrow, Celtic Field, Charcoal kiln	Veluwe, The Netherlands (2200 km ²)	0.5 m DTM; 6–10 points/m ²	Faster Region-based CNN (WODAN 2.0)	F1-score 70%
[85]	Maya Settlements (Aguada, Building, Platform)	Campeche, Mexico (230 km ²)	0.5 m DEM; 14.7 points/m ² (ground)	CNN (VGG-19)	Accuracy 95%
[84]	Bomb Crater, Charcoal Kiln, Barrow	Harz Mountains, Germany (47,000 km ²)	0.5 m DTM	CNN (Deeplab v3+)	IoU 76.8%
[67]	Burial Mounds	Romania (200 km ²)	0.5 m DEM; 2–6 points/m ²	ML (Random Forest)	Accuracy 96%

Table 1. Cont.

Authors	Archaeological Sites/Objects	Study's Location (Extent)	LiDAR Derivative and Resolution	Detection Method (Architecture/Algorithm)	Quality Evaluation
[71]	House, Wall, Pyramid, etc.	Mexico (-)	0.3 m DEM	CNN (VGG)	Precision 97%
[14]	Historic Mining Pits	Dartmoor National Park, UK (-)	0.25 m and 0.5 m DSM	TL CNN (DeepMoon)	Recall 80% (0.5 m DSM) and 83% (0.25 m DSM)
[81]	Prehistoric Roundhouses, Shieling Huts, Clearance Cairns	Arran, Scotland (432 km ²)	0.25 m DTM; 2.75 points/m ² (ground)	TL CNN (ResNet-18)	Accuracy (Roundhouse 73%, Huts 26%, Cairns 20%)
[7]	Burial Mounds	Brittany, France (246.7 km ²)	0.25 m DTM; 14 points/m ²	ML (Random Forest)	-
(b)					
[29]	Hillforts	England (130,000 km ²), Alto Minho, Portugal (2220 km ²), Galicia, Spain (30,000 km ²)	1 m DTM; 0.5 and 2 points/m ²	CMX (Semantic Segmentation)	F1-score 66%
[26]	Maya Structures	Tabasco, Mexico (885 km ²), Petén, Guatemala (615 km ²)	1 m DEM; 2.07 points/m ² (ground)	CNN (YOLOv3)	F1-score 80%
[18]	Burial Mounds	Alto Minho, Portugal (2220 km ²)	1 m DTM	Region-based CNN (YOLOv3)	Detection Rate 72.53%
[79]	Stone Walls	Northeastern CT, USA (-)	1 m DEM	CNN (U-Net)	Recall 89% Precision 93% F1-score 91%
[42]	Shipwreck	Alaska, and Puerto Rico, USA (-)	1 m DEM	TL CNN (YOLOv3)	F1-score 92%
[62]	Stone Wall, Pottery	Chun Castle, UK (-)	1 m DSM	ML (Support Vector Machine)	Accuracy >70%
[78]	Burial Mounds	Galicia, Spain (29,574 km ²)	1 m DTM	Region-based CNN (YOLOv3)	Detection Rate 89.5%, Precision 66.75%
[10]	Shell Rings	South Carolina, USA (6712 km ²)	1.5 m DEM	Mask Region-based CNN	Accuracy ~75%
[54]	Field Systems (Medieval Terraced Slopes, and Ridges and Furrows)	Southern Vosges, France (1462 km ²)	1 m DEM; 5 points/m ²	ML (Random Forest) and DL (Fully Connected Networks)	F-score 64–91% (ML) and 55–77% (DL)
[68]	Relict Charcoal Hearths	New England, USA (493 km ²)	1 m DEM; 2 points/m ²	CNN (U-Net)	F1-score 86%

Table 1. Cont.

Authors	Archaeological Sites/Objects	Study's Location (Extent)	LiDAR Derivative and Resolution	Detection Method (Architecture/ Algorithm)	Quality Evaluation
[24]	Maya Structures	Petén, Guatemala (2144 km ²)	1 m DEM	Mask Region-based CNN (U-Net)	Accuracy 95%
[11]	Viking Age Fortress	Bornholm, Denmark (42,036 km ²)	1.6 m DTM	ML (Random Forest)	-
[74]	Hollow Way, Stream, Pathway, Lake, Street, Ditch, etc.	Lower Saxony, Germany (-)	1 m DTM	Hierarchical CNN	Accuracy 91%
[24]	Maori Storage Pits	New Zealand (-)	1 m DEM	ML (Template Matching)	-
(c)					
[12]	Historical Terrain Anomalies	Eifel Region, Germany (0.01 km ²)	DTM; 200–300 points/m ²	ML (Support Vector Machine)	Recall 76–80% Precision 55–72% F1-score 57–81%
[2]	Clearance Cairns	Söderåsen, Sweden (-)	DTM; 0.5–1 points/m ²	CNN (U-Net segmentation)	Dice coefficient 84%
[64]	Archaeological Topography	Perticara, Italy (106.45 km ²)	DEM; 142 points/m ²	ML (Unsupervised ISODATA)	-
[6]	Earthwork Sites (Pit, Terrace, Sod Wall, Ditch)	Northland, New Zealand (-)	Low-quality DEM	Faster Region-based CNN (ResNet-101)	-
[75]	Barrow, Celtic Field, Charcoal Kiln	Veluwe, The Netherlands (2200 km ²)	DTM; 6–10 points/m ²	CNN (YOLOv4)	Precision 64%, F1-score 76%
[20]	Barrows and Celtic Fields	Veluwe, The Netherlands (440 km ²)	LiDAR images; 6–10 points/m ²	Region-based CNN (WODAN)	F1-score~70%
[4]	Barrow, Celtic Field, Charcoal Kiln	Veluwe, The Netherlands (437.5 km ²)	LiDAR images; 6–10 points/m ²	CNN (WODAN)	-
[69]	Grave, Mound, Pitfall Trap, Charcoal Burning Pit, Charcoal Kiln	Oppland, Norway (29 km ²)	-	ML (Template Matching)	-

Building on the capabilities of RS data, ML aims to train algorithms to learn from these data to make predictions or decisions. It has shown potential in automating archaeological feature detection, such as the use of RF algorithms for the detection of burial mounds in France and Spain [78], Viking ring fortresses throughout Denmark [11], ancient canals in Belize, Central America [25], and the use of CNN to detect ancient Maya structures in Guatemala [24]. These advancements hold significant potential for revolutionizing archaeological research by offering efficient, accurate, and scalable feature detection and analysis methods. Various studies have leveraged airborne LiDAR data and ML techniques to identify ancient structures and landscapes. For instance, ancient Chinese city walls were delineated using SS applied to DEMs derived from ALS data [8]. Similarly, other

studies have employed CNNs to detect walls and houses from derivatives of noisy airborne LiDAR data, with applications that extend to mapping ancient walls in different countries [6,17,61,62,71,79]. Additionally, deep semantic models have been proposed for predicting the locations of ancient agricultural terraces and walls, highlighting the potential of cost-effective raster data in transforming archaeological research [19,54,60,62,80]. Such studies provide valuable references for ancient site detection and monitoring, offering insights into cultural heritage preservation and aiding in reconstructing urban structures and their functions. Detection of burial mounds has been another focus, with ML and DL methods applied to elevation models derived from airborne LiDAR data [7,13,18,41,67,78]. Similarly, segmentation models trained from scratch have been used to detect clearance cairns in forested areas, enhancing understanding of historical agricultural activity and settlement organization [2]. Innovative approaches employing ML-based detection have also been applied to Celtic fields, barrows, and charcoal kilns in airborne LiDAR data, showcasing the potential of automatic measures for archaeological research evaluation [6,22,23,61,62,64,70,79–82]. Moreover, mapping Maya archaeological sites has been difficult due to their location in dense forests and rugged landscapes. Combining LiDAR data and CNNs can make it easier and more efficient to analyze these sites [25,26,85]. Also, terrain and topographical features, such as hillforts, have been the focus of various studies [12,29,64] employing different ML methods. Despite challenges such as the complexity of LiDAR data, these studies demonstrate encouraging potential, with proposed models freely available for other users to adapt to their needs. In addition, TL has also been employed to reduce the cost and hazards of underwater archaeology using bathymetric LiDAR data [42]. Techniques like Mask R-CNN and segmentation have also been utilized to detect relict charcoal hearths and kilns, achieving impressive results in object detection and instance segmentation [13,22,84]. Other innovative approaches, such as TL for detecting historic mining pits, have shown strong potential for broader archaeological tasks, which can demonstrate efficient semi-automated object detection and can distinguish between natural and manmade features [14]. Similarly, ML approaches have been employed to detect Viking ring fortresses in Denmark and hollow roads using DL and image processing methods [11,53,74]. Furthermore, DL and airborne LiDAR derivatives have been studied to detect archaeological shell rings, providing insights into native inhabitants and their socioeconomic networks [10].

Table 1a–c highlights the lack of consistency in evaluation metrics across studies. Various metrics such as accuracy, precision, recall, F1-score, IoU, and detection rate have been computed, which makes direct comparisons between studies difficult. Definitions and contextual explanations of these evaluation metrics are provided in Appendix A (Glossary). This inconsistency can limit the ability to assess how well ML methods are performing on a global scale and across different archaeological contexts. Calculating a unified and consistent metric would enable a standardized comparison, providing deeper insights into the effectiveness and reliability of these methods in archaeological applications. Another insight is that CNN-based methods, especially with high-resolution airborne LiDAR derivatives and for localized studies with detailed archaeological features (for instance, the Netherlands and Mexico [15,71,82,85]), demonstrate superior detection accuracy and precision compared to traditional ML methods. While higher resolution LiDAR tends to achieve better detection performances, as seen in cases like tar production kilns [5] and burial mounds [67], coarser resolutions are used for larger study extents but often result in moderate detection rates. CNN-based methods are the most commonly used, demonstrating versatility by detecting a wide range of archaeological features, including walls [6,17,61,71,79], mounds [13,18,41,78], and Maya structures [24,26,85]. However, ML methods are effective for more straightforward feature types (like canals and linear structures) but less robust for complex features. Techniques such as filtering and applying

enhancement methods on LiDAR data help overcome challenges like vegetation obstruction and erosion effects [64]. While specific LiDAR guidelines are still evolving (both for data acquisition and for data validation), expertise in landscape analysis remains crucial for accurate assessments, with LiDAR technology enriching our understanding of landscapes over time [47]. In summary, integrating airborne LiDAR derivatives with ML techniques appears to offer promising avenues for supporting archaeological research and cultural heritage preservation, with various studies showcasing the effectiveness of these approaches across diverse archaeological tasks. Furthermore, collaboration between archaeologists and ML experts may contribute to the refinement of detection methods, and adopting standard evaluation measures can facilitate cross-study comparisons, fostering the development of human-centered ML methods for archaeological feature detection. Finally, archaeological sites remain vulnerable to a range of threats, including environmental factors (such as natural disasters or climate change) and human activities (such as urban development or looting). Despite these challenges, the mentioned studies provide meaningful insights aiding the interpretation of archaeological sites and planning management strategies to protect and preserve them.

6. Discussion

6.1. *The Value of LiDAR and Machine Learning Integration*

LiDAR is particularly valuable in archaeological research because of its ability to penetrate dense vegetation, such as forest canopies, and capture detailed measurements of the ground surface. This capability is crucial for discovering unknown archaeological features in heavily vegetated areas, which are often challenging to investigate through a traditional framework or optical imagery. LiDAR data are often processed into DTMs or other visualization products like hillshades or LRMs, allowing archaeologists to visualize topographical changes that may indicate the presence of archaeological remains such as earthworks, mounds, or ditches. The increasing availability of LiDAR datasets reinforces the potential for automated analysis.

The application of ML/DL to derivatives of aerial LiDAR data represents a transformative shift in archaeology, especially for subtle feature detection over expansive areas. Traditional archaeological methods, including manual analysis of remotely sensed data, are often time-consuming and labor-intensive. Therefore, the integration of ML/DL techniques offers promising solutions to automate the detection process, enhancing efficiency and potentially reducing costs associated with extensive manual surveys. DL models, particularly CNNs, have demonstrated state-of-the-art performance in object recognition tasks and are well suited for analyzing raster images derived from aerial LiDAR. Various ML/DL approaches, such as object detection, segmentation, and classification, have been applied to identify diverse archaeological features. Examples include the detection of burial mounds using DL or RF [7,18,47,67,78,86], qanat shaft using CNNs [1], tar production kilns using a U-Net-based algorithm [5], hollow roads [53], stone walls and farmsteads [55,79], and hillforts [29,37]. These automated methods can assist archaeologists in quickly identifying potential areas of interest across large regions, which then can guide subsequent fieldwork and reduce the need for exhaustive manual analysis.

Understanding this interdisciplinary domain requires knowledge of LiDAR's ability to produce high-resolution DTMs or DEMs by penetrating vegetation, overcoming limitations of traditional methods. Historically, archaeologists have relied on the manual analysis and visual interpretation of remotely sensed data. However, this approach faces severe limitations in the age of big data. Manual analysis is limited by scalability, subjectivity, and human processing capacity. They are inefficient and often impractical for large-scale systematic mapping [2,4,12,14,15,19,74] and can be influenced by an individual's experience,

knowledge, and potential biases, which can cause inconsistent results and difficulties in reproducibility [5,12,24,60]. Finally, the dimensionality and resolution of LiDAR derivatives often exceed the processing capacity of the human visual system, making it challenging to consistently identify subtle or complex features [4,14]. ML/DL models address these through automated feature learning, large-scale data processing, enhanced detection, reproducibility and consistency, semantic and instance segmentation, and use of LiDAR-derived visualization techniques (e.g., LRM, hillshade, openness). Unlike traditional image processing techniques, ML/DL techniques can generalize from given examples rather than requiring human-defined parameters, making them robust to variations in archaeological object morphology and context [15]. They can rapidly process vast amounts of LiDAR data that would be infeasible through manual methods alone [2,5,14,15]. This significantly reduces the time and labor investment for localization tasks, freeing up expert time for analysis, field validation, and interpretation. They can also identify subtle depressions and patterns in the landscape that might be missed by the human eye, especially for ephemeral or partially preserved archaeological features [11,14,29,72], and can reduce the variability inherent in manual interpretations [5,22].

Moreover, ML/DL methods can provide detailed information about archaeological exact boundaries, size, and coverage [24,41,72,75], which are useful for features with irregular shapes or those covering extensive areas, where rectangular bounding boxes might be insufficient [75]. Instance segmentation, for instance, applying Mask R-CNN, further distinguishes individual instances of objects, even if they overlap [24,85]. To improve ML/DL results, different visualization techniques of LiDAR derivatives can be used to make archaeological features more visible to both human interpreters and ML/DL algorithms, thereby improving model performance [26]. Moreover, traditional ML evaluation metrics (Precision, Recall, F1-score, IoU) are being adapted or new ones developed (e.g., centroid-based, pixel-based measures) to better align with archaeological needs and interpretation, recognizing that archaeological objects have distinctive shapes and geospatial relevance often not captured by generic computer vision metrics [26,82].

DL, particularly using NNs like Faster R-CNN, Mask R-CNN, and U-Net variants, generally outperforms traditional ML in detecting complex archaeological features, especially when image-based data are involved [54,61,74]. For barrows and burial mounds, DL models like WODAN2.0 and Mask R-CNN achieved F1-scores up to 0.70 and 0.82, respectively [15,75], while ML methods like RF reached even higher scores (e.g., 0.98 for Neolithic burial mounds) [7,20]. Citizen scientists outperformed WODAN2.0 on Celtic fields [15]. This might be because human experts recognize the checkerboard patterns of Celtic fields, which have few natural parallels, whereas the object detection model focuses on individual plots, a shape more abundant in the landscape. WODAN2.0 underperformed on charcoal kilns (F1-score = 18%) due to shape diversity and limited training data [15], though Mask R-CNN improvements showed F1-scores up to 0.84 [22,75]. For linear features like hollow roads, CarcassonNet achieved an F1-score of 0.50 [53]. Moreover, FCNs slightly outperformed RF for ridges and furrows [54]. DL also excelled in detecting stone walls (FCN with a U-Net-like architecture and F1-score = 0.88) [17,60], mining pits (F1-score of up to 0.87) [14], and shipwrecks (YOLOv3 F1-score = 0.92), often aided by TL [42]. For canals and wetland fields, ML methods like RF were effective but prone to FP, and the Simple Local Relief Model (SLRM) proved to be the most useful visualization for revealing these canals [25]. In detecting hillforts, early CNNs had many FPs, but CMX improved the results (F1-score = 66% in Iberia) [29]. CNNs are overwhelmingly preferred over recent alternatives like vision transformers or hybrid CNN-transformer models in archaeological detection. This preference is primarily due to CNNs' established performance, architectural

suitability for image and spatial data [15,53,75], and adaptability to data scarcity, which are critical factors in archaeological contexts [1,2,41,79,85].

6.2. Current Applications and Achievements

The effectiveness of ML/DL methods in archaeology is context-dependent, influenced by the morphology of features [14,20,24,41,53,78], their visibility and preservation [3,20,24,33,78], data type and quality [10,27,29,34], and the availability of training data [3–5,15,57]. Features with regular shapes like burial mounds are easier to detect [1,14,68], while irregular or eroded features pose challenges [10,20,63]. Furthermore, resolution and point density critically influence the model performance, especially for small or subtle features [5,20,27]. DL models like CNNs, Mask R-CNN, YOLO, and U-Net excel at detecting consistent shapes like circular features and known architectural elements with pixel-level precision but need large datasets and domain-informed training. For instance, Mask R-CNN was successfully applied to detect charcoal fireplaces, achieving 83% recall and 87% accuracy [1]. YOLOv3 demonstrated high precision (97%) and acceptable recall (64%) for detecting archaeological tumuli [78]. YOLOv8 segmented Maya archaeological platforms and annular structures, achieving an IoU of 84% and 81%, respectively [76]. Similarly, the DeepMoon model, a CNN-based TL approach, showed promise for detecting historic mining pits and other circular height changes [14]. U-Net has been effective in detecting tar production kilns in boreal forests [5]. Moreover, RF handles imbalanced data well and is useful for metric-based [54,67] and agricultural [25,54] features. It can also be used in hybrid approaches, for instance, to classify multispectral satellite data to delineate areas where DL models should focus, thereby reducing FP [54]. For example, RF was used to verify Mask R-CNN results for shell ring locations by analyzing multispectral and SAR data [10]. It is known for its robustness to data distributions, relatively few parameters to optimize, and good performance in routine RS applications [7,25,54,67]. SVMs, on the other hand, are suited for anomaly detection with limited data but high risk of FP [12]. They are ideal for terrain anomalies like foxholes [12] and linear features like stone walls or ditches that remain [17,62]. OBIA and template matching are often handcrafted and rely on explicit prior knowledge of object properties [4,20], and they work for simple shape objects like circles or rectangles, and well-defined shapes, but lack generalizability [4,12,50,68]. Multimodal approaches and human–AI collaboration are vital for refining results, validating detections, and overcoming challenges such as data heterogeneity, biases, and small datasets, pointing to the need for adaptive, transferable workflows and archaeologically meaningful evaluation metrics.

The strategic choices in ML/DL architectures, the management of essential hyperparameters, and the practical limitations encountered are critical to understanding the current state and future potential of this field. At the core of many archaeological DL applications are CNNs, which are image feature extractors and classifiers [15]. A key architecture family is the R-CNN series, designed for object detection, meaning that they both localize and classify objects within a larger image [20]. Specific architectures of this family are Faster R-CNN and Mask R-CNN. Faster R-CNN was used in workflows like WODAN2.0 to improve detection speed through a region proposal network [15]. Furthermore, Mask R-CNN adds instance segmentation capabilities by extending Faster R-CNN and adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition. It enables instance segmentation and pixel-level delineation of objects [24,72,75]. This is particularly valuable for complex and irregularly shaped archaeological features and effectively is applied to relict charcoal hearths [22,29,32,75]. Like the R-CNN family, YOLO frameworks (YOLOv3–v8) are also widely adopted due to their speed and accuracy as one-stage detectors [26,42,78], with YOLOv4, YOLOv5, and YOLOv8 tailored for LiDAR-

based archaeological tasks such as burial mound detection [18,75,76] and YOLOv8-CDD for underwater cultural heritage detection [77].

Beyond object detection, SS models like U-Net are used for pixel-level classification in detecting features like stone walls, terraces, and kilns [5,17,41,61,68,80]. It can classify each pixel of an image with a corresponding class, providing precise boundaries of objects [24,61]. Moreover, pre-trained CNNs such as VGG-16, ResNet, AlexNet, and support TL for small archaeological datasets, since they are often pre-trained on large generic datasets like ImageNet and allow leveraging features learned from vast, diverse image collections and then fine-tune them for specific archaeological tasks, reducing training time and increasing generalization ability [4,15,26,54,74]. For instance, a model pre-trained on Lunar LiDAR data was successfully adapted to detect historic mining pits, hypothesizing applicability to any circular height change features [14,54]. Lastly, hybrid and fusion models are emerging to address the complexity of archaeological detection by combining different data types or multiple DL models. This includes early fusion techniques that integrate LiDAR-derived LRMs with aerial orthoimages [29], or the fusion of different SS models like DeepLabv3+ and U-Net for ancient agricultural terrace detection [80]. These multimodal approaches aim to emulate the multi-data analysis performed by expert archaeologists, providing richer contextual information [29].

6.3. Technical and Practical Challenges and Generalization Issues

Despite the significant potential and early successes, the application of ML/DL to LiDAR derivatives for archaeological detection faces several critical challenges that need careful consideration and ongoing research. The first challenge is the availability and preparation of training data. ML and DL techniques heavily depend on labeled data, requiring large volumes of annotated datasets [10,41,54,74,84] that are scarce, costly, and prone to bias and uncertainty [8,18,24,54,75]. TL and data augmentation are common strategies to mitigate this limitation [10,54,72,74]. Moreover, archaeological datasets often exhibit severe class imbalance [20,41,54,80], which, if not properly addressed, can cause models to bias predictions towards the majority class and result in deceptive accuracy metrics [26] unless mitigated with techniques like downsampling the majority class, oversampling the minority class, and using specialized loss functions [41,53,54]. Furthermore, training DL models demands substantial computational resources, including powerful GPUs (Graphics Processing Units) and significant memory [7,10,29,61,80]. This can make exhaustive hyperparameter tuning impractical within typical project constraints [17]. Insufficient interpretation has been shown to lead to limitations in training experiments [40]. Moreover, analyzing archaeological features in LiDAR derivatives is a task that requires archaeological expertise. This makes the creation of large, well-labeled datasets expensive and time-consuming. Considering that, the quality and consistency of expert interpretation can vary, which will potentially introduce bias and error into the training data. This can degrade classifier accuracy. The entire process, from data acquisition to interpretation, involves assumptions and decisions by the operator, which can introduce subjectivity and compromise validity if not properly reported. This underscores the crucial need for standardized documentation, including metadata (data about data) and paradata (documentation of process), to ensure scientific transparency, replicability, and reflexivity.

Beyond training data volume, challenges also exist in working with LiDAR data of varying point densities and intrinsic precision. Particularly, detecting features in low-density data is difficult. While higher point density LiDAR coverages may become available in the future, researchers currently face issues with available data quality and quantity. Another issue is the lack of publicly available archaeological data due to ethical concerns regarding site protection. Standardized, open-access, and large datasets are lacking, which

makes the evaluation and comparison of the performance of different ML/DL models across various archaeological contexts difficult. Fairly comparing different detection methods is challenging because their performance heavily depends on the datasets, metrics, and evaluation methods used. This highlights the need for adopting standard evaluation measures within the archaeological community. There is also a gap in providing efficient and automatic data structuring pipelines for existing datasets that were not originally acquired for heritage detection purposes.

DL models are often referred to as black boxes because of their complex and multi-layered structures, which make it difficult to understand precisely how these models arrive at their predictions [7,24,54]. This opacity can lead to distrust from non-expert users and challenges in identifying the causes of errors or refining the model [22,81]. Moreover, models often struggle to generalize across different regions or data formats (e.g., three-channel RGB images vs. single-channel LiDAR DTMs) [26,29,81], and fine-tuning models with small, targeted datasets from new regions is often necessary [29]. Differences in data formats can make TL challenging, as most existing pretrained models are designed for RGB imagery [14,54], and optimal parameter settings or visualization choices found for one specific task or class may not be optimal for others [19]. Another issue is the high rates of FP and FN, which are common due to similarities between archaeological and natural (e.g., drift-sand dunes, rock outcrops, natural depressions) or modern (e.g., roundabouts, house roofs) features [15,18,29,78]. Archaeological features, especially those that are subtle or degraded, often have similar morphologies to natural or artificial non-archaeological shapes in the landscape. This bird's-eye perspective challenge in LiDAR data means that objects with similar forms (e.g., small mounds, pits) can be difficult for ML models to distinguish without additional context or validation [18,29,37,47]. While post-processing validation steps, such as analyzing the 3D shape of potential detections, can help reduce FP, a persistently high rate of FP can require significant effort for subsequent ground truthing. Conversely, FN is also a concern. There is often a trade-off between precision (minimizing FP) and recall (minimizing FN) [15]. For field archaeologists, high precision is crucial due to limited time and financial resources for investigation, while for cultural heritage managers, high recall is paramount for effective conservation and protection [12,19,29] and ensuring that no actual features are overlooked. Further investigation can later confirm or reject the findings.

The variability in archaeological feature characteristics and landscape contexts, such as the ambiguous boundaries of ancient features [80], presents further challenges and requires further dedicated studies. Archaeological remains are in diverse sizes, shapes, and levels of preservation. Moreover, their appearance in LiDAR data can be influenced by factors like erosion, vegetation density, and the specific LiDAR processing techniques used. Developing models that can effectively detect this wide range of features across topographically varied landscapes is complex. Model transferability between different geographic regions and archaeological sites is not guaranteed and often requires fine-tuning or retraining [36,40]. The necessary resolution and point density of LiDAR data also vary depending on the size and detail of the archaeological features being sought. Lower densities might miss smaller or less distinct features. Moreover, deriving suitable raster products (like DTMs or other visualizations) from raw aerial LiDAR involves numerous decisions and algorithms. Therefore, the choice of processing steps can significantly impact the performance of ML/DL models and, as a result, the visibility of archaeological features. Documenting this complex workflow is crucial for scientific transparency and replicability, which is not yet standardized.

Although deep CNN models hold significant potential, they are still not commonly used in detecting archaeological remains [31]. To our knowledge, there has been limited

evaluation of CNNs' object-segmentation capabilities. Most CNN-based object detection techniques in this domain rely on two-stage detectors, such as R-CNNs, Faster R-CNNs, and Mask R-CNNs. While these approaches are robust and often highly accurate, they can face challenges related to slower processing speeds than one-stage detectors. Two-stage CNN detectors first generate region proposals (i.e., areas in the image that might contain objects) and then classify each proposed region and refine its bounding box, while one-stage detectors predict object locations and classes directly in a single step, making them faster but sometimes less accurate [4,65,74]. Additionally, difficulties persist in selecting suitable DL approaches, generating training datasets, and accurately labeling data. Furthermore, adaptable ML methods applied for the segmentation of unstructured 3D data are still under discussion and are not yet consolidated [40]. Three notable contributions include the multi-level multi-resolution SS approach, which utilizes RF algorithms for classification [86], the implementation of the PointConv architecture for high-accuracy classification of 3D point clouds [3], and the application of DL on 3D airborne LiDAR data for SS and object detection of historical defensive architectures [40]. These approaches can offer innovative solutions for analyzing LiDAR-generated datasets and detecting archaeological structures with improved accuracy and efficiency. Direct processing of raw 3D point cloud data preserves full geometric detail [9,65,87], enabling high-accuracy detection and classification of subtle and irregular archaeological features without the information loss of 2D or voxel conversions [3,25,73], and supports advanced modeling, segmentation, and multispectral analysis [34]. However, most existing DL approaches for point clouds are primarily focused on other fields, like robotics, autonomous driving, and indoor modeling [40]. Adapting these methods for use in cultural heritage and landscape contexts requires significant modifications. Model generalization and adaptability are, in fact, challenging when applying systems developed in one region or for one type of feature to areas with different site typologies and landscapes, and normally require fine-tuning [29]. Models often lack generalizability, struggle with similar-looking non-archaeological features, and require expert input, while also lacking transparency and semantic richness crucial for archaeological interpretation [9,34,65,70].

6.4. Practical Considerations and Implementation Needs

Integrating information from various sensors, such as combining airborne LiDAR data with photogrammetric data [31], aerial imagery, multispectral/hyperspectral imaging, geophysical surveys, and ground-based LiDAR scans [87] or utilizing multispectral LiDAR [25,87], offers the opportunity to overcome the limitations of single data sources by extracting richer and more detailed information about the potential archaeological features. However, challenges remain in effectively integrating these diverse datasets, considering variations in resolution, penetration, texture, color, accuracy, and the dynamic nature of the environment. Therefore, these techniques are not commonly used, and there is limited evidence of their effective detection of hidden remains. Moreover, the development of hybrid ML/DL models that fuse predictions from different models, combine DL with traditional methods, or integrate ML/DL with external knowledge sources or processes [20,80] could further enhance the accuracy and interpretability of archaeological feature detection algorithms. The application of these hybrid approaches can lead to high detection and segmentation performance even with relatively small training datasets. It could address a common limitation in archaeological contexts where large training datasets are scarce. By combining the strengths of different methods, these hybrid models can potentially achieve higher accuracy and reduce FP compared to using a single method. However, challenges exist. Implementing such combined approaches can require substantial computational resources and processing time [29]. While hybrid detection methods can be

fast, the process often requires significant human expertise for creating and refining training datasets, validating results, and interpreting findings, especially when dealing with complex or heterogeneous archaeological features [5]. Also, managing various data types introduces storage challenges. Important practical aspects could be excluding negative zones (i.e., areas where archaeological information cannot be obtained, such as built-up areas or areas with insufficient data quality) [37] and establishing specific arrangements for the long-term storage and archiving of digital data products, considering the necessary storage space [5]. Therefore, their successful implementation needs careful consideration of data requirements, computational resources, and the critical role of human expertise in the workflow.

Furthermore, the successful implementation of ML/DL in archaeology requires close interdisciplinary collaboration and knowledge integration between archaeologists, computer scientists, and RS experts. Applying complex computer algorithms remains uncommon for many archaeologists, because it often requires the expertise of computer science specialists. Additionally, a barrier of meaning [33] exists, which represents the gap between the expert archaeologist's knowledge and the knowledge learned by the machine. One of the most prevalent examples is the machine's difficulty in differentiating archaeological objects from natural or modern landscape elements that share a similar visual morphology in RS data [20,26,75]. For instance, human interpreters can easily distinguish a roundabout from a barrow, even though both might appear as circular, positive elevations in LiDAR data [20,75]. In contrast, automated detection approaches primarily rely on perception from small segments of an image, typically from a single data source, lacking the broader comprehension that humans apply [20]. To address this, it is fundamental to enhance the involvement of archaeologists in the learning process. This enables them to contribute their expertise and to provide domain knowledge to the machine [82]. Archaeologists provide the expertise necessary for identifying and interpreting potential features, defining target classes for ML/DL models, and validating results through ground truthing. Computer scientists develop and refine the ML/DL algorithms. And, RS specialists handle data acquisition and processing. Integrating archaeological knowledge into the ML/DL workflow, such as using LBR or incorporating specific archaeological object patterns, can improve model performance and reduce FP [16,18,47]. Moreover, international collaboration and the establishment of standardized datasets [82] are essential for facilitating the evaluation of ML/DL models in feature detection and classification within the archaeological field. By leveraging technological innovations and fostering collaboration among archaeological teams, we can accelerate the pace and improve the quality of archaeological investigations, ultimately contributing to a deeper understanding of ancient civilizations and the preservation of cultural heritage.

6.5. Opportunities and Future Research Directions

6.5.1. Improving Precision and Reducing False Positives

Future research directions may focus on addressing the identified challenges. Finding ways to deal with the lack of labeled training data is very important. This could be achieved using methods like active learning, weakly supervised learning, or more effective data augmentation techniques such as generating synthetic point cloud data and 3D bounding box labels [40]. Furthermore, creating and sharing standardized, open-access datasets with high-quality annotations and ground truth validation would significantly facilitate the development and comparison of ML/DL models. Moreover, further investigation is needed to understand. How do different LiDAR processing techniques impact the visibility of various archaeological features? How to optimize these techniques for automated detection?

6.5.2. Multi-Modal Data Fusion and Complementary Data Integration

Exploring the fusion of LiDAR with other RS data, such as multispectral imagery or photogrammetry, and exploring the potential opportunities of hyperspectral LiDAR may provide additional information to improve detection accuracy and reduce FP. Therefore, future work should include ablation studies to quantitatively assess how much each individual data source (or the combination of them) contributes to the model's performance. This helps prove whether using multiple data sources actually improves the results. Likewise, refining post-processing validation methods such as using 3D information from the point cloud or incorporating spatial context analysis could help filter out non-archaeological features.

6.5.3. Advanced DL Architectures and Techniques

Future works could also apply semi-supervised and self-supervised algorithms for archaeological detection. These approaches aim to leverage large volumes of unlabeled data through unsupervised pretraining, followed by supervised fine-tuning with limited labeled data, to improve prediction accuracy [34,65,84]. To further evaluate their potential, case studies should be investigated where these models have been applied to multi-modal datasets (e.g., combining LiDAR with optical imagery) and assess how this impacts performance and generalizability across diverse archaeological landscapes.

6.5.4. Enhancing Training Data and Labeling Methodologies

Developing transferable methodological approaches that can adapt to varying primary data densities, particularly addressing low-density applications, is needed. It is also important to develop more robust models that can deal with the inherent differences in archaeological features, can work well across various landscapes and data types, are able to leverage higher point density LiDAR data as it becomes available, and can incorporate new data sources like LiDAR intensity values.

6.5.5. Integration into Archaeological Workflows and Decision Support Systems

Finally, fostering interdisciplinary and international collaboration and knowledge integration is crucial. Collaboration among surveyors, archaeologists, and software engineers for method development is encouraged. Moreover, future research should also place site findings within their broader regional and temporal context [32] and consider investigating archaeological features that extend over national borders by combining LiDAR data and AI.

7. Conclusions

Our work serves as a valuable bridge between traditionally separate disciplines by demonstrating how AI-driven object detection using CNNs and LiDAR can be effectively applied to archaeological research. By clearly mentioning underutilized technical operations, such as direct 3D point cloud analysis, broad-scale model generalization, and human–AI collaborative workflow, this study promotes meaningful collaboration across archaeology, geosciences/RS, computer science/engineering, heritage management, and public engagement. It can support the growing interdisciplinary momentum in landscape archaeology and can contribute to the development of sustainable, scalable, and scientifically robust methodologies for both academic research and heritage practices.

Integrating airborne LiDAR derivatives with ML techniques represents a notable advancement in archaeological research. The combination of LiDAR's high-resolution terrain mapping capabilities with the automation capabilities of ML algorithms has the potential to enhance current methodologies, enabling more efficient and systematic detection of hidden archaeological landscapes and structures.

Through this literature review, we have explored the diverse applications of ML-based approaches in archaeological feature detection, ranging from identifying ancient settlements to detecting burial mounds and urban complexes. These studies demonstrate the potential of ML techniques, particularly DL models, in augmenting traditional archaeological methods and facilitating a deeper understanding of past civilizations.

Despite notable progress and the demonstrated potential of ML-based approaches across diverse archaeological tasks, several critical challenges persist that shape future research directions. A major hurdle for DL models is their reliance on large volumes of high-quality, labeled training data, which are currently scarce, costly, and prone to bias in archaeology. Future research must focus on strategies to address this scarcity, including the creation and sharing of standardized, open-access datasets with high-quality annotations and ground truth validation. This would greatly facilitate the development and comparative evaluation of ML/DL models. Given the limitations of labeled data, increasing the use of techniques like TL and weakly supervised learning is vital. TL has shown promise in improving model accuracy and reducing training time and data requirements in archaeological contexts. Exploring more effective data augmentation techniques, such as generating synthetic point cloud data, also offers potential solutions. Moreover, the current lack of consistency in evaluation metrics across studies makes direct comparisons challenging and limits the assessment of ML method performance on a global scale. Establishing standardized evaluation measures and benchmark tasks is crucial for enabling consistent comparison, providing deeper insights into method effectiveness and reliability, and developing human-centered ML methods.

Last but not least, the successful implementation of ML/DL in archaeology demands close interdisciplinary collaboration and knowledge integration among archaeologists, computer scientists, and RS experts. Overcoming the barrier of meaning requires enhancing archaeologists' involvement in the learning process, allowing them to contribute domain expertise and validate results. Fostering such collaborations is fundamental to developing sustainable, scalable, and scientifically robust methodologies. Continued exploration and innovation may help deepen our understanding of ancient civilizations, support cultural heritage preservation, and inspire future generations to engage with the rich diversity of human history.

Author Contributions: Conceptualization, G.B., R.Z.; methodology, G.B., E.M. and R.Z.; investigation, R.Z.; resources, G.B.; writing—original draft preparation, R.Z.; writing—review and editing, G.B., E.M. and R.Z.; supervision, G.B., E.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ALS	Airborne Laser Scanning
ANNs	Artificial Neural Networks
BIM	Building Information Model
CNN	Convolutional Neural Network
DEMs	Digital Elevation Models
DL	Deep Learning

DNNs	Deep Neural Networks
DTMs	Digital Terrain Models
DSMs	Digital Surface Models
FN	False Negative
FP	False Positive
GIS	Geographic Information Systems
GNSS	Global Navigation Satellite System
IMU	Inertial Measurement Unit
LBR	Location-Based Ranking
LiDAR	Light Detection and Ranging
LRMs	Local Relief Models
ML	Machine Learning
NIR	Near-Infrared Range
OBIA	Object-Based Image Analysis
R-CNN	Mask Region-based Convolutional Neural Network
ReLU	Rectified Linear Units
RF	Random Forest
RNN	Recurrent Neural Network
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SS	Semantic Segmentation
SVM	Support Vector Machine
TL	Transfer Learning
TN	True Negative
TP	True Positive
UAV	Unmanned Aerial Vehicle

Appendix A Glossary

In archaeological remote sensing, various metrics are employed to evaluate the performance of ML and DL models.

Metric	Definition	Importance and Use Cases
Accuracy (Overall Accuracy)	Represents the ratio of correctly predicted instances to the total number of instances in a dataset. It is calculated as the number of overall correctly classified items (TP + TN) compared to the entire number of samples [26,42,53,60].	<ul style="list-style-type: none"> - Provides an initial general evaluation of model performance across all classes [85]. - Can be a poor or deceptive metric when dealing with imbalanced datasets (This is because a high accuracy might simply reflect the correct classification of the majority class (e.g., background), even if the minority class (e.g., archaeological objects) is poorly detected) [18,26,53].
Recall (Completeness, Detection Rate, Sensitivity)	Measures how many relevant objects are selected by the model. It is calculated as the number of TP (TP) divided by the sum of TP and FN [15,20].	<ul style="list-style-type: none"> - Indicates the model's ability to find all positive samples [26,42] (In archaeology, it is often more important to detect as many potential anomalies as possible, even if it means a higher rate of FP, to ensure no actual sites are missed during initial automated surveys.) - Higher recall is preferred when the cost of missing an object is high [12,42,71].

Metric	Definition	Importance and Use Cases
Precision (Correctness)	Measures how many of the selected items are relevant. It is calculated as the number of TP (TP) divided by the sum of TP and FP [15,20,53].	<ul style="list-style-type: none"> - Indicates the proportion of the model's predictions that are correct [42]. - It is influenced by the number of FP [26]. - High precision is desired when minimizing false alarms is critical [42].
F1-score (Dice–Sorensen coefficient)	The harmonic average of precision and recall. It is calculated as $2 \times (\text{Recall} \times \text{Precision}) / (\text{Recall} + \text{Precision})$ [15,20,53,68].	<ul style="list-style-type: none"> - Provides a balanced measure of the model's performance per class (especially useful when datasets have an uneven class distribution) [15,20,53]. - It offers a single value that considers both recall and precision, providing a more comprehensive view of performance than either metric alone [14,26,42].
Reliability	Refers to the consistency and trustworthiness of a measurement, method, or system [32,37], and contributes directly to the accuracy of results [87].	<ul style="list-style-type: none"> - Reliability is crucial for ensuring the scientific validity and practical applicability of research and automated tools, especially in fields like archaeology [37]
Intersection over Union (IoU, Jaccard index)	Quantifies the overlap between a predicted bounding box or segmentation mask and the ground truth (actual object) [72]. It is calculated as the area of intersection divided by the area of union between the predicted and ground truth regions [8,24,80].	<ul style="list-style-type: none"> - It is one of the most commonly used measures for assessing the performance of object detection methods [74]. - For SS, it measures the overlap between generated and ground truth masks at a pixel level [41]. - For object detection, a common threshold (e.g., 0.5) is applied to determine if a prediction counts as a True Positive [72,75]. - It is also effective in overcoming class imbalance issues in pixel-based classification [61]. - For discrete archaeological objects, it may not be ideal as it emphasizes the extent of overlap rather than the precise geographical location (centroid) of the object, which archaeologists often prioritize [74].
Mean Average Precision	It is typically defined as the average of the Average Precision (AP) values calculated for each object class, often at a specific IoU threshold (e.g., mAP@0.5 IoU). AP itself represents the area under the precision-recall curve for a given class [72,75].	<ul style="list-style-type: none"> - A popular metric for object detection models [72]. - Provides a single, comprehensive metric for evaluating object detection performance across multiple classes, considering both localization and classification accuracy [72,75]. - It allows for direct comparison of different object detection models.
Matthews Correlation Coefficient (MCC)	A balanced measure of the quality of binary classifications that can be used even if the classes are of very different sizes [24]. It takes into account all four values in a confusion matrix: TP, TN, FP, and FN [53].	<ul style="list-style-type: none"> - It is considered a more reliable indicator of model quality than accuracy or F1-score when datasets are imbalanced, as it accounts for the proportion of all four categories equally [24].

Metric	Definition	Importance and Use Cases
Receiver Operating Characteristic (ROC) Curve	It plots the True Positive Rate (recall) against the False Positive Rate for different classification thresholds [42].	<ul style="list-style-type: none"> - It summarizes the overall predictive ability of a model, with values between 0.7 and 0.8 being acceptable, above 0.8 excellent, and above 0.9 outstanding [42]. - It is less sensitive to class imbalance than overall accuracy [42].
Precision–Recall Curve	A graph that plots precision values against corresponding recall values for different classification thresholds [42].	<ul style="list-style-type: none"> - Particularly informative for evaluating models on imbalanced datasets, as it focuses on the performance of the positive class and provides insights into the trade-off between precision and recall [42].
Kappa Coefficient (Cohen’s Kappa)	Measures the agreement between the predicted classification and the true values, accounting for the possibility of agreement occurring by chance [60,62].	<ul style="list-style-type: none"> - Used in accuracy assessment for classification results [60,62] - Provides a more robust measure of agreement than simple accuracy, especially when evaluating agreement between a classified map and reference data [60,61].
Loss Functions (e.g., Mean Squared Error (MSE), Binary Cross-Entropy, Jaccard Loss, Categorical Focal Loss)	<p>A loss function is a mathematical function that quantifies the error or discrepancy between the predicted output of a model and the true (ground truth) labels [8,22,80].</p> <p>During the training process, the neural network parameters are iteratively adjusted to minimize this loss function [8,13,84].</p>	<ul style="list-style-type: none"> - Fundamental for guiding the learning process of ML/DL models [8,13,84]. - MSE is commonly used in regression tasks and for autoencoders [84]. - Binary Cross-Entropy is often used for binary classification problems [29,79]. - Jaccard Loss (based on the IoU) and Categorical Focal Loss are used to address class imbalance problems, especially in image segmentation tasks where a small number of pixels represent the target class [41,61].

References

- Jamil, A.H.; Yakub, F.; Azizan, A.; Roslan, S.A.; Zaki, S.A.; Ahmad, S.A. A Review on Deep Learning Application for Detection of Archaeological Structures. *J. Adv. Res. Appl. Sci. Eng. Technol.* **2022**, *26*, 7–14. [\[CrossRef\]](#)
- Küçükdemirci, M.; Landeschi, G.; Ohlsson, M.; Dell’Unto, N. Investigating Ancient Agricultural Field Systems in Sweden from Airborne LIDAR Data by Using Convolutional Neural Network. *Archaeol. Prospect.* **2023**, *30*, 209–219. [\[CrossRef\]](#)
- Richards-Rissetto, H.; Newton, D.; Al Zadjali, A. A 3D Point Cloud Deep Learning Approach Using Lidar to Identify Ancient Maya Archaeological Sites. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *VIII-M-1*, 133–139. [\[CrossRef\]](#)
- Lambers, K.; Verschoof-van der Vaart, W.B.; Bourgeois, Q.P. Integrating Remote Sensing, Machine Learning, and Citizen Science in Dutch Archaeological Prospection. *Remote Sens.* **2019**, *11*, 794. [\[CrossRef\]](#)
- Anttiroiko, N.; Groesz, F.J.; Ikäheimo, J.; Kelloniemi, A.; Nurmi, R.; Rostad, S.; Seitsonen, O. Detecting the Archaeological Traces of Tar Production Kilns in the Northern Boreal Forests Based on Airborne Laser Scanning and Deep Learning. *Remote Sens.* **2023**, *15*, 1799. [\[CrossRef\]](#)
- Bickler, S.H.; Jones, B. Scaling up Deep Learning to Identify Earthwork Sites in Te Tai Tokerau, Northland, New Zealand. *Archaeol. N. Z.* **2021**, *64*, 16–24.
- Guyot, A.; Hubert-Moy, L.; Lorho, T. Detecting Neolithic Burial Mounds from LiDAR-Derived Elevation Data Using a Multi-Scale Approach and Machine Learning Techniques. *Remote Sens.* **2018**, *10*, 225. [\[CrossRef\]](#)
- Wang, S.; Hu, Q.; Wang, S.; Ai, M.; Zhao, P. Archaeological Site Segmentation of Ancient City Walls Based on Deep Learning and LiDAR Remote Sensing. *J. Cult. Herit.* **2024**, *66*, 117–131. [\[CrossRef\]](#)
- Pierdicca, R.; Paolanti, M.; Matrone, F.; Martini, M.; Morbidoni, C.; Malinverni, E.S.; Frontoni, E.; Lingua, A.M. Point Cloud Semantic Segmentation Using a Deep Learning Framework for Cultural Heritage. *Remote Sens.* **2020**, *12*, 1005. [\[CrossRef\]](#)
- Davis, D.S.; Caspari, G.; Lipo, C.P.; Sanger, M.C. Deep Learning Reveals Extent of Archaic Native American Shell-Ring Building Practices. *J. Archaeol. Sci.* **2021**, *132*, 105433. [\[CrossRef\]](#)

11. Stott, D.; Kristiansen, S.M.; Sindbæk, S.M. Searching for Viking Age Fortresses with Automatic Landscape Classification and Feature Detection. *Remote Sens.* **2019**, *11*, 1881. [[CrossRef](#)]
12. Storch, M.; De Lange, N.; Jarmer, T.; Waske, B. Detecting Historical Terrain Anomalies With UAV-LiDAR Data Using Spline-Approximation and Support Vector Machines. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 3158–3173. [[CrossRef](#)]
13. Trier, Ø.D.; Reksten, J.H.; Løseth, K. Automated Mapping of Cultural Heritage in Norway from Airborne Lidar Data Using Faster R-CNN. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *95*, 102241. [[CrossRef](#)]
14. Gallwey, J.; Eyre, M.; Tonkins, M.; Coggan, J. Bringing Lunar LiDAR Back down to Earth: Mapping Our Industrial Heritage through Deep Transfer Learning. *Remote Sens.* **2019**, *11*, 1994. [[CrossRef](#)]
15. Verschoof-van der Vaart, W.B.; Lambers, K.; Kowalczyk, W.; Bourgeois, Q.P. Combining Deep Learning and Location-Based Ranking for Large-Scale Archaeological Prospection of LiDAR Data from the Netherlands. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 293. [[CrossRef](#)]
16. Wang, S.; Hu, Q.; Wang, F.; Ai, M.; Zhong, R. A Microtopographic Feature Analysis-Based LiDAR Data Processing Approach for the Identification of Chu Tombs. *Remote Sens.* **2017**, *9*, 880. [[CrossRef](#)]
17. Trotter, E.F.L.; Fernandes, A.C.M.; Fibæk, C.S.; Keßler, C. Machine Learning for Automatic Detection of Historic Stone Walls Using LiDAR Data. *Int. J. Remote Sens.* **2022**, *43*, 2185–2211. [[CrossRef](#)]
18. Canedo, D.; Fonte, J.; Seco, L.G.; Vázquez, M.; Dias, R.; Pereiro, T.D.; Hipólito, J.; Menéndez-Marsh, F.; Georgieva, P.; Neves, A.J.R. Uncovering Archaeological Sites in Airborne LiDAR Data With Data-Centric Artificial Intelligence. *IEEE Access* **2023**, *11*, 65608–65619. [[CrossRef](#)]
19. Laplaige, C.; Ramel, J.-Y.; Rodier, X.; Bai, S.; Guillaume, R. Extraction of Linear Structures from LIDAR Images Using a Machine Learning Approach. In Proceedings of the IMEKO International Conference on Metrology for Archaeology and Cultural Heritage, Torino, Italy, 19–21 October 2016; pp. 83–88.
20. Verschoof-Van der Vaart, W.B.; Lambers, K. Learning to Look at LiDAR: The Use of R-CNN in the Automated Detection of Archaeological Objects in LiDAR Data from the Netherlands. *J. Comput. Appl. Archaeol.* **2019**, *2*, 31–40. [[CrossRef](#)]
21. Blau, S. Archaeology: Definition. In *Encyclopedia of Global Archaeology*; Smith, C., Ed.; Springer: New York, NY, USA, 2014; p. 449. ISBN 978-1-4419-0465-2.
22. Bonhage, A.; Eltaher, M.; Raab, T.; Breuß, M.; Raab, A.; Schneider, A. 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. *Archaeol. Prospect.* **2021**, *28*, 177–186. [[CrossRef](#)]
23. Berganzo-Besga, I.; Orengo, H.A.; Canela, J.; Belarte, M.C. Potential of Multitemporal Lidar for the Detection of Subtle Archaeological Features under Perennial Dense Forest. *Land* **2022**, *11*, 1964. [[CrossRef](#)]
24. Bundzel, M.; Jaščur, M.; Kováč, M.; Lieskovský, T.; Sinčák, P.; Tkáčik, T. Semantic Segmentation of Airborne Lidar Data in Maya Archaeology. *Remote Sens.* **2020**, *12*, 3685. [[CrossRef](#)]
25. Doyle, C.; Luzzadder-Beach, S.; Beach, T. Advances in Remote Sensing of the Early Anthropocene in Tropical Wetlands: From Biplanes to Lidar and Machine Learning. *Prog. Phys. Geogr. Earth Environ.* **2023**, *47*, 293–312. [[CrossRef](#)]
26. Character, L.; Beach, T.; Inomata, T.; Garrison, T.G.; Luzzadder-Beach, S.; Baldwin, J.D.; Cambranes, R.; Pinzón, F.; Ranchos, J.L. Broad-scale Deep Learning Model for Archaeological Feature Detection across the Maya Area. *J. Archaeol. Sci.* **2024**, *169*, 106022. [[CrossRef](#)]
27. Jones, B.; Bickler, S.H. High Resolution LiDAR Data for Landscape Archaeology in New Zealand. *Archaeology* **2017**, *60*, 35–44.
28. Cigna, F.; Balz, T.; Tapete, D.; Caspari, G.; Fu, B.; Abballe, M.; Jiang, H. Exploiting satellite SAR for archaeological prospection and heritage site protection. *Geo-Spat. Inf. Sci.* **2024**, *27*, 526–551. [[CrossRef](#)]
29. Canedo, D.; Fonte, J.; Dias, R.; do Pereiro, T.; Gonçalves-Seco, L.; Vázquez, M.; Georgieva, P.; Neves, A.J.R. Automated Detection of Hillforts in Remote Sensing Imagery With Deep Multimodal Segmentation. *Archaeol. Prospect.* **2024**, *32*, 297–311. [[CrossRef](#)]
30. Lasaponara, R.; Masini, N. Satellite Synthetic Aperture Radar in Archaeology and Cultural Landscape: An Overview. *Archaeol. Prospect.* **2013**, *20*, 71–78. [[CrossRef](#)]
31. Kadhim, I.; Abed, F.M. A Critical Review of Remote Sensing Approaches and Deep Learning Techniques in Archaeology. *Sensors* **2023**, *23*, 2918. [[CrossRef](#)]
32. Ikäheimo, J. Detecting Pitfall Systems in the Suomenselkä Watershed, Finland, with Airborne Laser Scanning and Artificial Intelligence. *J. Archaeol. Sci. Rep.* **2023**, *51*, 104216. [[CrossRef](#)]
33. Argyrou, A.; Agapiou, A. A Review of Artificial Intelligence and Remote Sensing for Archaeological Research. *Remote Sens.* **2022**, *14*, 6000. [[CrossRef](#)]
34. Yang, S.; Hou, M.; Li, S. Three-Dimensional Point Cloud Semantic Segmentation for Cultural Heritage: A Comprehensive Review. *Remote Sens.* **2023**, *15*, 548. [[CrossRef](#)]
35. Adamopoulos, E.; Rinaudo, F. UAS-Based Archaeological Remote Sensing: Review, Meta-Analysis and State-of-the-Art. *Drones* **2020**, *4*, 46. [[CrossRef](#)]

36. Vinci, G.; Vanzani, F.; Fontana, A.; Campana, S. LiDAR Applications in Archaeology: A Systematic Review. *Archaeol. Prospect.* **2024**, *32*, 81–101. [CrossRef]
37. Lozić, E.; Štular, B. Documentation of Archaeology-Specific Workflow for Airborne LiDAR Data Processing. *Geosciences* **2021**, *11*, 26. [CrossRef]
38. Masini, N.; Lasaponara, R. Airborne Lidar in Archaeology: Overview and a Case Study. In Proceedings of the Computational Science and Its Applications—ICCSA, Ho Chi Minh City, Vietnam, 24–27 June 2013; Murgante, B., Misra, S., Carlini, M., Torre, C.M., Nguyen, H.-Q., Taniar, D., Apduhan, B.O., Gervasi, O., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; Volume 7972, pp. 663–676.
39. Holden, N.; Horne, P.; Bewley, R. High-Resolution Digital Airborne Mapping and Archaeology. *Nato Sci. Ser. I Life Behav. Sci.* **2002**, *337*, 173–180.
40. Cappellazzo, M.; Patrucco, G.; Spanò, A. ML Approaches for the Study of Significant Heritage Contexts: An Application on Coastal Landscapes in Sardinia. *Heritage* **2024**, *7*, 5521–5546. [CrossRef]
41. Banasiak, P.Z.; Berezowski, P.L.; Zapłata, R.; Mielcarek, M.; Duraj, K.; Stereńczak, K. Semantic Segmentation (U-Net) of Archaeological Features in Airborne Laser Scanning—Example of the Białowieża Forest. *Remote Sens.* **2022**, *14*, 995. [CrossRef]
42. Character, L.; Ortiz, A., Jr.; Beach, T.; Luzzadder-Beach, S. Archaeologic Machine Learning for Shipwreck Detection Using Lidar and Sonar. *Remote Sens.* **2021**, *13*, 1759. [CrossRef]
43. Davis, D.S.; Cook Hale, J.W.; Hale, N.L.; Johnston, T.Z.; Sanger, M.C. Bathymetric LiDAR and Semi-Automated Feature Extraction Assist Underwater Archaeological Surveys. *Archaeol. Prospect.* **2024**, *31*, 171–186. [CrossRef]
44. Devereux, B.J.; Amable, G.S.; Crow, P.; Cliff, A.D. The Potential of Airborne Lidar for Detection of Archaeological Features under Woodland Canopies. *Antiquity* **2005**, *79*, 648–660. [CrossRef]
45. Zhang, F.; Hassanzadeh, A.; Kikkert, J.; Pethybridge, S.J.; van Aardt, J. Comparison of UAS-Based Structure-from-Motion and LiDAR for Structural Characterization of Short Broadacre Crops. *Remote Sens.* **2021**, *13*, 3975. [CrossRef]
46. Crabb, N.; Carey, C.; Howard, A.J.; Brolly, M. Lidar Visualization Techniques for the Construction of Geoarchaeological Deposit Models: An Overview and Evaluation in Alluvial Environments. *Geoarchaeology* **2023**, *38*, 420–444. [CrossRef]
47. Collaro, C.; Herkommer, M. Research, Application, and Innovation of LiDAR Technology in Spatial Archeology. In *Encyclopedia of Information Science and Technology*, 6th ed.; IGI Global: Pennsylvania, PA, USA, 2025; pp. 1–33. ISBN 978-1-6684-7366-5.
48. Simbolon, R.S.; Comer, A. Unveiling the Past: LiDAR Technology’s Role in Discovering Hidden Archaeological Sites. *J. Ilmu Pendidik. Dan Hum.* **2023**, *12*, 14–30. [CrossRef]
49. 3. On the Archaeology of Imaginary Media. In *Media Archaeology*; Huhtamo, E., Parikka, J., Eds.; University of California Press: Berkeley, CA, USA, 2019; pp. 48–69. ISBN 978-0-520-94851-8.
50. Davis, D.S. Object-based Image Analysis: A Review of Developments and Future Directions of Automated Feature Detection in Landscape Archaeology. *Archaeol. Prospect.* **2019**, *26*, 155–163. [CrossRef]
51. Evans, D.H.; Fletcher, R.J.; Pottier, C.; Chevance, J.-B.; Soutif, D.; Tan, B.S.; Im, S.; Ea, D.; Tin, T.; Kim, S.; et al. Uncovering Archaeological Landscapes at Angkor Using Lidar. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 12595–12600. [CrossRef]
52. Evans, D.; Hanus, K.; Fletcher, R. The Story beneath the Canopy: An Airborne LiDAR Survey over Angkor, Phnom Kulen and Koh Ker, Northwestern Cambodia. In *Across Space and Time: Papers from the 41st Conference on Computer Applications and Quantitative Methods in Archaeology, Perth, 25–28 March 2013*; Traviglia, A., Ed.; Amsterdam University Press: Amsterdam, The Netherlands, 2015; pp. 36–44.
53. Verschoof-van der Vaart, W.B.; Landauer, J. Using CarcassonNet to Automatically Detect and Trace Hollow Roads in LiDAR Data from the Netherlands. *J. Cult. Herit.* **2021**, *47*, 143–154. [CrossRef]
54. Herrault, P.-A.; Poterek, Q.; Keller, B.; Schwartz, D.; Ertlen, D. Automated Detection of Former Field Systems from Airborne Laser Scanning Data: A New Approach for Historical Ecology. *Int. J. Appl. Earth Obs. Geoinform.* **2021**, *104*, 102563. [CrossRef]
55. Johnson, K.M.; Ouimet, W.B. Rediscovering the Lost Archaeological Landscape of Southern New England Using Airborne Light Detection and Ranging (LiDAR). *J. Archaeol. Sci.* **2014**, *43*, 9–20. [CrossRef]
56. Verhoeven, G.J. Are We There yet? A Review and Assessment of Archaeological Passive Airborne Optical Imaging Approaches in the Light of Landscape Archaeology. *Geosciences* **2017**, *7*, 86. [CrossRef]
57. Bickler, S.H. Machine Learning Arrives in Archaeology. *Adv. Archaeol. Pract.* **2021**, *9*, 186–191. [CrossRef]
58. Fiorucci, M.; Khoroshiltseva, M.; Pontil, M.; Traviglia, A.; Del Bue, A.; James, S. Machine Learning for Cultural Heritage: A Survey. *Pattern Recognit. Lett.* **2020**, *133*, 102–108. [CrossRef]
59. LaRue, K. Digital Archaeology: Detection of Archaeological Structures Using Convolutional Neural Networks on Aerial LiDAR Data. *WWU Honors College Senior Projects*. 2023, p. 639. Available online: https://cedar.wvu.edu/wwu_honors/639 (accessed on 7 September 2024).
60. Mohleli, M.G.; Adam, E.; Schoeman, M.H. The potential for LiDAR using support vector machine (SVM) to detect archaeological stone-walled structures in Khutwaneng, Bokoni. *S. Afr. Archaeol. Bull.* **2023**, *78*, 33–42.

61. Tiwari, A.; Silver, M.; Karnieli, A. A Deep Learning Approach for Automatic Identification of Ancient Agricultural Water Harvesting Systems. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *118*, 103270. [[CrossRef](#)] [[PubMed](#)]
62. Kadhim, I.; Abed, F.M. The Potential of LiDAR and UAV-Photogrammetric Data Analysis to Interpret Archaeological Sites: A Case Study of Chun Castle in South-West England. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 41. [[CrossRef](#)]
63. Karamitrou, A.; Sturt, F.; Bogiatzis, P.; Beresford-Jones, D. Towards the Use of Artificial Intelligence Deep Learning Networks for Detection of Archaeological Sites. *Surf. Topogr. Metrol. Prop.* **2022**, *10*, 044001. [[CrossRef](#)]
64. Masini, N.; Abate, N.; Gizzi, F.T.; Vitale, V.; Minervino Amodio, A.; Sileo, M.; Biscione, M.; Lasaponara, R.; Bentivenga, M.; Cavalcante, F. UAV LiDAR Based Approach for the Detection and Interpretation of Archaeological Micro Topography under Canopy—The Rediscovery of Peticara (Basilicata, Italy). *Remote Sens.* **2022**, *14*, 6074. [[CrossRef](#)]
65. Guo, Y.; Wang, H.; Hu, Q.; Liu, H.; Liu, L.; Bennamoun, M. Deep Learning for 3d Point Clouds: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 4338–4364. [[CrossRef](#)]
66. Berganzo Besga, I. New Computational Methods for Automated Large-Scale Archaeological Site Detection. Ph.D. Thesis, Universitat Rovira i Virgili, Tarragona, Spain, 2023.
67. Niculiță, M. Geomorphometric Methods for Burial Mound Recognition and Extraction from High-Resolution LiDAR DEMs. *Sensors* **2020**, *20*, 1192. [[CrossRef](#)]
68. Suh, J.W.; Anderson, E.; Ouimet, W.; Johnson, K.M.; Witharana, C. Mapping Relict Charcoal Hearths in New England Using Deep Convolutional Neural Networks and Lidar Data. *Remote Sens.* **2021**, *13*, 4630. [[CrossRef](#)]
69. Kermit, M.; Reksten, J.H.; Trier, Ø.D. *Towards a National Infrastructure for Semi-Automatic Mapping of Cultural Heritage in Norway*; Archaeopress: Oxford, UK, 2018; pp. 159–172.
70. Farmakis, I.; DiFrancesco, P.-M.; Hutchinson, D.J.; Vlachopoulos, N. Rockfall Detection Using LiDAR and Deep Learning. *Eng. Geol.* **2022**, *309*, 106836. [[CrossRef](#)]
71. Albrecht, C.M.; Fisher, C.; Freitag, M.; Hamann, H.F.; Pankanti, S.; Pezzutti, F.; Rossi, F. Learning and Recognizing Archeological Features from LiDAR Data. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 5630–5636.
72. Guyot, A.; Lennon, M.; Lorho, T.; Hubert-Moy, L. Combined Detection and Segmentation of Archeological Structures from LiDAR Data Using a Deep Learning Approach. *J. Comput. Appl. Archaeol.* **2021**, *4*, 1–19. [[CrossRef](#)]
73. Lyons, M.; Fecher, F.; Reindel, M. From LiDAR to Deep Learning: A Case Study of Computer-Assisted Approaches to the Archaeology of Guadalupe and Northeast Honduras. *IT-Inf. Technol.* **2022**, *64*, 233–246. [[CrossRef](#)]
74. Kazimi, B.; Thiemann, F.; Malek, K.; Sester, M.; Khoshelham, K. Deep Learning for Archaeological Object Detection in Airborne Laser Scanning Data. In Proceedings of the 2nd Workshop on Computing Techniques for Spatio-Temporal Data in Archaeology and Cultural Heritage, Melbourne, Australia, 25–28 August 2018; Volume 15, pp. 21–35.
75. Olivier, M.; Verschoof-van der Vaart, W. Implementing State-of-the-Art Deep Learning Approaches for Archaeological Object Detection in Remotely-Sensed Data: The Results of Cross-Domain Collaboration. *J. Comput. Appl. Archaeol.* **2021**, *4*, 274–289. [[CrossRef](#)]
76. Zhang, J.; Ringle, W.; Willis, A.R. Unveiling ancient Maya settlements using aerial LiDAR image segmentation. *arXiv* **2024**, arXiv:2403.05773. [[CrossRef](#)]
77. Liu, T. YOLOv8-CDD: A Salient Target Detection Model for Underwater Cultural Heritage in Complex Environments. *Int. Core J. Eng.* **2025**, *11*, 332–342.
78. Berganzo-Besga, I.; Orenge, H.A.; Lumbreras, F.; Carrero-Pazos, M.; Fonte, J.; Vilas-Estévez, B. Hybrid MSRM-Based Deep Learning and Multitemporal Sentinel 2-Based Machine Learning Algorithm Detects near 10k Archaeological Tumuli in North-Western Iberia. *Remote Sens.* **2021**, *13*, 4181. [[CrossRef](#)]
79. Suh, J.W.; Ouimet, W.B.; Dow, S. Reconstructing and Identifying Historic Land Use in Northeastern United States Using Anthropogenic Landforms and Deep Learning. *Appl. Geogr.* **2023**, *161*, 103121. [[CrossRef](#)]
80. Wang, Y.; Liu, C.; Tiwari, A.; Silver, M.; Karnieli, A.; Zhu, X.X.; Albrecht, C.M. Deep Semantic Model Fusion for Ancient Agricultural Terrace Detection. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 17–20 December 2022; pp. 4888–4892.
81. Trier, Ø.D.; Cowley, D.C.; Waldeland, A.U. Using Deep Neural Networks on Airborne Laser Scanning Data: Results from a Case Study of Semi-Automatic Mapping of Archaeological Topography on Arran, Scotland. *Archaeol. Prospect.* **2019**, *26*, 165–175. [[CrossRef](#)]
82. Fiorucci, M.; Verschoof-Van Der Vaart, W.B.; Soleni, P.; Le Saux, B.; Traviglia, A. Deep Learning for Archaeological Object Detection on LiDAR: New Evaluation Measures and Insights. *Remote Sens.* **2022**, *14*, 1694. [[CrossRef](#)]
83. Schneider, A. Looking at LiDAR Pixel-by-Pixel: A Critical Approach. *arXiv* **2024**, arXiv:2409.11532v1.
84. Kazimi, B.; Thiemann, F.; Sester, M. Semi Supervised Learning for Archaeological Object Detection in Digital Terrain Models. In *International Conference on Cultural Heritage and New Technologies*; Propylaeum: Heidelberg, Germany, 2020; pp. 219–225.

85. Somrak, M.; Džeroski, S.; Kokalj, Ž. Learning to Classify Structures in ALS-Derived Visualizations of Ancient Maya Settlements with CNN. *Remote Sens.* **2020**, *12*, 2215. [[CrossRef](#)]
86. Mazzacca, G.; Grilli, E.; Cirigliano, G.; Remondino, F.; Campana, S. Seeing among Foliage with LIDAR and Machine Learning: Towards a Transferable Archaeological Pipeline. *International Archives of the Photogrammetry. Remote Sens. Spat. Inf. Sci.* **2022**, *46*, 365–372.
87. Takhtkeshha, N.; Mandlbürger, G.; Remondino, F.; Hyypä, J. Multispectral Light Detection and Ranging Technology and Applications: A Review. *Sensors* **2024**, *24*, 1669. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.