Automated detection of former field systems from airborne laser scanning data: a new approach for Historical Ecology

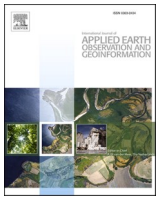
Pierre-Alexis Herrault, Quentin Poterek, Benjamin Keller, Dominique Schwartz, Damien Ertlen

International Journal of Applied Earth Observations and Geoinformation 104 (2021) 102563

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Automated detection of former field systems from airborne laser scanning data: a new approach for Historical Ecology

P.-A. Herrault *, Q. Poterek, B. Keller, D. Schwartz, D. Ertlen

UMR 7362 LIVE CNRS, University of Strasbourg, 3, rue de l'Argonne, 67000 Strasbourg, France

ARTICLE INFO

Keywords:
Remote Sensing
Former Field Systems
Airborne Laser Scanning
Deep Learning
Local Binary Patterns
Historical Ecology

ABSTRACT

Former field systems (FFS) are the most widespread traces of ancient activities in present European landscapes and represent significant perturbations to ecosystems. Through its ability to penetrate forest canopies and detect microlandforms, Airborne Laser Scanning data reveal archaeological relics over large areas, from periods older than the first available Historical Topographic Maps. Mapping these traces from ALS-derived data (e.g. Digital Elevation Model (DEM)) thus allows for a determination of a new temporal baseline in order to evaluate the effects of a longer history on current patterns of biodiversity. Here, we evaluate the ability of traditional machine learning (Random Forest-RF) and deep learning (Fully Connected Networks-FCN) models to detect Medieval Terraced slopes and Ridges and Furrows (RaF) from an ALS-derived DEM in the southern Vosges (1462 km^2). We used a combination of Local Binary Patterns and topographical metrics to measure properties of FFS and to train detection models. We then assessed the relative performance of each model semantically and spatially. Our results demonstrated the high suitability of our approach for reproducing major trends in the landscape with a high level of similarity between the predicted and reference spatial patterns (Structural Similarity Index - SSIM > 0.75). RF outperformed FCN for Terraced Slopes, whilst minimizing the false positive rate. FCN slightly outperformed RF for the RaF dataset but showed promising abilities to survey unseen data with a low sensitivity to annotation errors. We suggest that this approach has the potential to offer new spatio-temporal possibilities in Historical Ecology studies as a means of automatically detecting past archaeological ecosystems from a landscape to a regional scale.

1. Introduction

Landscape history provides a potential avenue for explaining some of the current patterns of biodiversity (Rhemtulla et al., 2009; Bürgi et al., 2015). Land cover and land use (LUCC) changes produce legacy effects on community structure, biomass, and regional C dynamics (Caspersen et al., 2000; Rhemtulla et al., 2009; Smithwick et al., 2007), which, in turn, alter processes such as soil nutrient dynamics and biodiversity over decades (Fraterrigo et al., 2005; Grossmann and Mladenoff, 2008), centuries (Compton and Boone, 2000), and even millennia (Dupouey et al., 2002). Former field systems (FFS) are the most widespread traces of ancient activities in current European landscapes, and they are the witnesses of deep mutations within ecosystems. Over the millennia, humans have modified the environment by digging, subdividing, and ploughing land for agricultural purposes (Kanianska, 2016). These persistent field systems can maintain themselves for many centuries as microlandforms. These features include ridges and furrows, terraced slopes, clearing cairns, and Celtic fields (see (Hesse, 2020 and Georges-Leroy, 2020) to get an overview). In Europe, these agriculture-related landforms started to appear in the Bronze Age (Froehlicher et al., 2016) and then became widely established during the later Middle Ages (Schwartz et al., 2020). Extreme historical events (e.g. the Thirty Years' War in Europe) and the emergence of new agricultural techniques led to the abandonment of these fields between the 17th and 19th centuries. From a morphological standpoint, these topographical features exhibit both irregular and regular shapes, depending on local topographic conditions and the degree of preservation. In general, FFS presently covered by forest or grasslands tend to be well preserved, whereas features covered by crops tend to be less visible.

Historical topographic and cadastral maps have been used successfully to assess the effects of past landscapes and their trajectories on the current biodiversity of grasslands (Adriaens et al., 2006; Cousins and Eriksson, 2001; Cousins et al., 2007; Cousins, 2009) and forests (Vellend et al., 2006; Piessens and Hermy, 2006; Kolk et al., 2017). Old aerial and

* Corresponding author.
E-mail address: pierre-alexis.herrault@live-cnrs.unistra.fr (P.-A. Herrault).

https://doi.org/10.1016/j.jag.2021.102563
Received 8 June 2021; Received in revised form 10 September 2021; Accepted 27 September 2021
Available online 12 October 2021

even satellite archives complete this cartographical information from the early 1930s to today (Krauss et al., 2010; Bommarco et al., 2014; Herrault et al., 2016). However, these sources only offer the possibility of reconstructing a relatively recent past, i.e. the last 200–300 years. Older landscape disturbances, such as medieval or earlier agricultural activities, potentially influence existing patterns of biodiversity (Compton and Boone, 2000; Dupouey et al., 2002), and this influence is not captured by the above-mentioned strategies. Moreover, these older landscapes can be masked by existing landscapes, making their detection from optical observation systems nearly impossible. Airborne Laser Scanning (ALS) has been proposed as a means of addressing these limitations, and serve as a tool for mapping archaeological landscapes and estimating ecological habitat continuity (Dupouey et al., 2002; Georges-Leroy and Montes, 2013). Through its ability to penetrate forest canopies and detect microlandforms, ALS allows for the discovery of new archaeological sites, and provides additional information for previously identified sites (for an extensive review, see (Parcak, 2009). ALS surveys have increased over the last decade, and multiple large-scale LiDAR products are often available, at no or low cost, through national mapping agencies. Nonetheless, the inherent attributes of ALS-derived elevation data, such as single-depth channels or changing-illumination properties, complicate the automated pattern recognition based on these data (Trier et al., 2019; Verschoof-van and Lambers, 2019). Consequently, maximising the amount of exploitable information from these data requires novel approaches.

Numerous automated approaches have been developed to facilitate the detection of archaeological landscapes from remotely sensed data. Approaches based on remote sensing, specifically using ALS data, can be divided into four categories (Verschoof-van and Lambers, 2019; Trier et al., 2019): template matching (De Boer, 2005; Schneider et al., 2015; Toumazet et al., 2017; Trier and Pilø, 2012; Trier et al., 2015), knowledge-based algorithms (Riley, 2009), object-based approaches (Davis et al., 2019; Freeland et al., 2016; Sevara et al., 2016; Kramer, 2015), and machine learning approaches. The latter approach performs very well when identifying various archaeological relics (Guyot et al., 2018; Menze et al., 2006; Menze and Ur, 2012; Trier et al., 2019; Trier et al., 2015; Verschoof-van and Lambers, 2019; Lambers et al., 2019). ML relies on the analysis of archaeological samples that are defined by a range of features, calculated using ancillary data obtained from digital elevation models, aerial photographs, and satellite images. Deep learning (DL), a subfield of the machine learning that relies on neural networks, has outperformed traditional ML methods on many applications (Verschoof-van and Lambers, 2019; Lambers et al., 2019; Trier et al., 2019). As computational resources have greatly improved over the last decade, this approach has been highly successful due to its higher predictive performances and diminishing processing times. Despite these recent advances, there are still uncertainties regarding the optimal methods for analysing ALS data at a large scale. This uncertainty is a major limitation regarding the use of ALS data within landscape ecology studies. First, DL is often favoured over traditional ML, although questions remain concerning the superiority of DL when using ALS-derived elevation data, most often contained in single-channel images. Second, earlier studies used to systematically employ DL models, such as Convolutional Neural Networks (CNNs), but the quality of the results proved to be spatially uneven depending on the surveyed landform (Verschoof-van and Lambers, 2019; Lambers et al., 2019; Trier et al., 2019; Albrecht et al., 2019).

So far, the detection of archaeological sites via ML has primarily relied on the use of the Random Forest (RF) algorithm, a nonlinear ensemble classifier (Guyot et al., 2018; Menze et al., 2006; Menze and Ur, 2012). RF models are trained using various features derived from visual interpretation techniques (VITs) or and/or morphological descriptors (Guyot et al., 2018; Toumazet et al., 2017). The number of hyperparameters to estimate is low, thereby making these models relatively easy to handle. Previous studies showed a strong performance (True-Positive Rate (TPR) >90%); however, these studies were limited because of the lack of generalisation in the learning decision trees when the output models were applied to independent data sets. RF can also be very sensitive to labelling errors, which may appear in archaeological training sets, although these errors are rarely addressed in archaeological surveys (Kramer, 2015).

CNN-based approaches have been recently used in archaeological studies. Nonetheless, they require a very large training set to perform well, which is sometimes difficult to achieve in archaeology. One of the strengths of CNNs is the availability of transfer learning oand domain adaptation techniques. These approaches require training a CNN using a generic image set, and then applying the model to a small data set (from a different domain (Wang and Deng, 2018)). As the pretrained parameters can be reused directly, the training time is reduced considerably, and the generalisation ability is increased. Transfer learning has been successfully used in archaeological surveys that rely on various image sources, including aerial photography and satellite-based remote sensing (Trier et al., 2019; Zingman et al., 2016; Gallwey et al., 2019). Despite these advances, there are several major limitations for these models beyond the commonly cited 'black box' issue (Verschoof-van and Lambers, 2019). First, DL models (including CNNs) are particularly sensitive to unbalanced data sets, a well-known issue in archaeological training sets. To overcome these limitations, researchers use data augmentation and apply transformations, such as flip or rotation in predefined directions (Trier et al., 2019; Verschoof-van and Lambers, 2019). Nonetheless, the results are mixed. A possible explanation is the difficulty in reproducing the finer-scale variations of landforms, or the frequently changing luminosity observed in data acquired under uncontrolled flying conditions, as is often the case with ALS data. The rebalancing of the data set is therefore ineffective and classification outputs are not satisfactory, namely for the under-represented class in the learning set. Moreover, questions also remain in terms of the initialisation process, because of the need to estimate a high number of parameters. Studies have shown that existing pretrained models, such as ResNet or VGG-16 (He et al., 2016), are less prone to overfitting and can generalise fairly well compared to models trained on smaller data sets. However, these models have been trained with three-channel RGB images, not with single-channel images, making the transfer learning with ALS data challenging. A recent deep CNN, trained using Lunar LiDAR data sets (a large scale digital DEM dataset acquired by the Lunar Reconnaissance Orbiter (Zuber et al., 2010)), successfully detected circular archaeological structures (Gallwey et al., 2019) from a DEM, with recall rates of 80%–83%. These models are designed to be highly receptive to elevation changes in single-channel DEM images but remain specific to circular objects and are transferable to only a few other forms. Finally, pretrained CNN models applied to image classification and object detection are usually trained on prior fixed input image size (i.e., from $224 \times 224 \times 3$ to approximately $512 \times 512 \times 3$). This may be particularly limiting for detecting archaeological objects, as they have attributes observable at various scales (i.e., from the local to the landscape scale), which are potentially very different from the input image size.

Here we propose an innovative approach for automatically detecting Middle Age terraced slopes (TS) and ridges and furrows (RaF) in the southern Vosges Mountains, France, within a 1462 km^2 survey region. We used interpolated ALS data, having a 1 m resolution, provided by the French National Mapping Agency (RGE IGN). We then trained a Fully Connected Network (FCN) with attributes obtained from the local binary patterns (LBP) to classify TS and RaF. LBP are texture operators that are invariant to rotation and illumination changes, a feature that is potentially well-suited to characterising FFS from ALS-derived elevation data. FCN models (a type of DL model) rely on precalculated features possibly based on various image sizes and therefore offer greater flexibility for extracting heterogeneous forms, such as FFS, at a large scale. We then compared our results with RF outputs to assess the respective performances of the ML and DL models and to evaluate our results with the published literature.

2. Material and Methods

2.1. Study Site

Our study area was constrained by the LiDAR data coverage (1462 km^2). The coverage incorporates the southern part of the Vosges Mountains in the Haut Rhin District (Fig. 1). The site is centred around the Hautes Vosges, the most elevated part of the Vosges Mountains. Elevations in the study area range from 135 to 1306 m asl. Various natural environments characterise the study site because of variations in geology and local climate. Winters are cold, and summers remain cool relative to the surrounding plains (July mean temperature of 11 °C vs 20 °C for the nearby Alsace plain) (François and Humbert, 2000). Precipitation may be substantial, commonly more than 2000 mm per year on ridges, in contrast to 540 mm per year in Colmar, 40 km to the west. Currently, forests cover 70% of the study area, but landscapes are also greatly structured by mountain farming and vineyards.

The benefits of working on the Vosges mountains for such a study are: (1) Human settlements in the Vosges mountains are ancient and well documented. First evidences about FFS go back to the 10th century; (2) a large amount of agricultural traces are visible in the landscapes (Froehlicher et al., 2016; Schwartz et al., 2020); (3) Vosges mountains have one of the largest and oldest forested area in France, and appear as a valuable site to evaluate the effects of past disturbances on the current biodiversity.

2.2. Terraced slopes and Ridges and Furrows

We investigated TS and RaF, as these landforms are the most common FFS remaining in the Vosges Mountains, and they have been evaluated in previous studies (Hesse, 2020; Kramer, 2015). The studied TS include lynchets and sloping agricultural terraces, which share similar morphological characteristics, although they differ in their origins. Lynchets are talus slopes formed at the top of hedgerows by field ploughing over an extended period. They may create extended parcelled areas with narrow plots running perpendicular to the slope (Schwartz et al., 2020). These plots and their dating remain poorly documented; however, some OSL dating suggests these features were installed during the Middle Ages (Froehlicher et al., 2016). In contrast, agricultural terraces are intentional breaks in the slope that are built to section a steep slope into workable plots and create favourable conditions for crops by limiting erosion. At present, TS are covered by forests or grasslands; however, they are not constrained by existing cadastral boundaries. RaF are archaeological patterns characterised by a succession of mounds and troughs. They are formed by a system of ploughing that was typical for the Middle Ages. As a consequence, the remaining RaF exhibit networks of rectangular parcels, irregular in width, and running parallel to each other. RaF may also be covered by forest, grasslands, and crops. Unlike TS, RaF boundaries may correspond to existing cadastral boundaries because they were abandoned more recently, and the remaining ridges can be used as actual boundaries to delimit new arable lands or grasslands (Fig. 2).

2.3. Data

We used ALS data obtained from the French Mapping Agency (RGE ALTI) under their Open Government License. The ALS data for the southern part of the Vosges Mountains were acquired in January 2013 and comprise 1 km by 1 km tiles of point cloud data (.las file v.1.2 compressed to laz). Each point was categorised into one of eight classes, including “ground”, “building”, and “vegetation”. The average ground point density per square meter was 5, although this varied strongly (1–7) among the tiles, depending on vegetation density and the presence of buildings. This dataset was available as Digital Elevation Model (DEM),

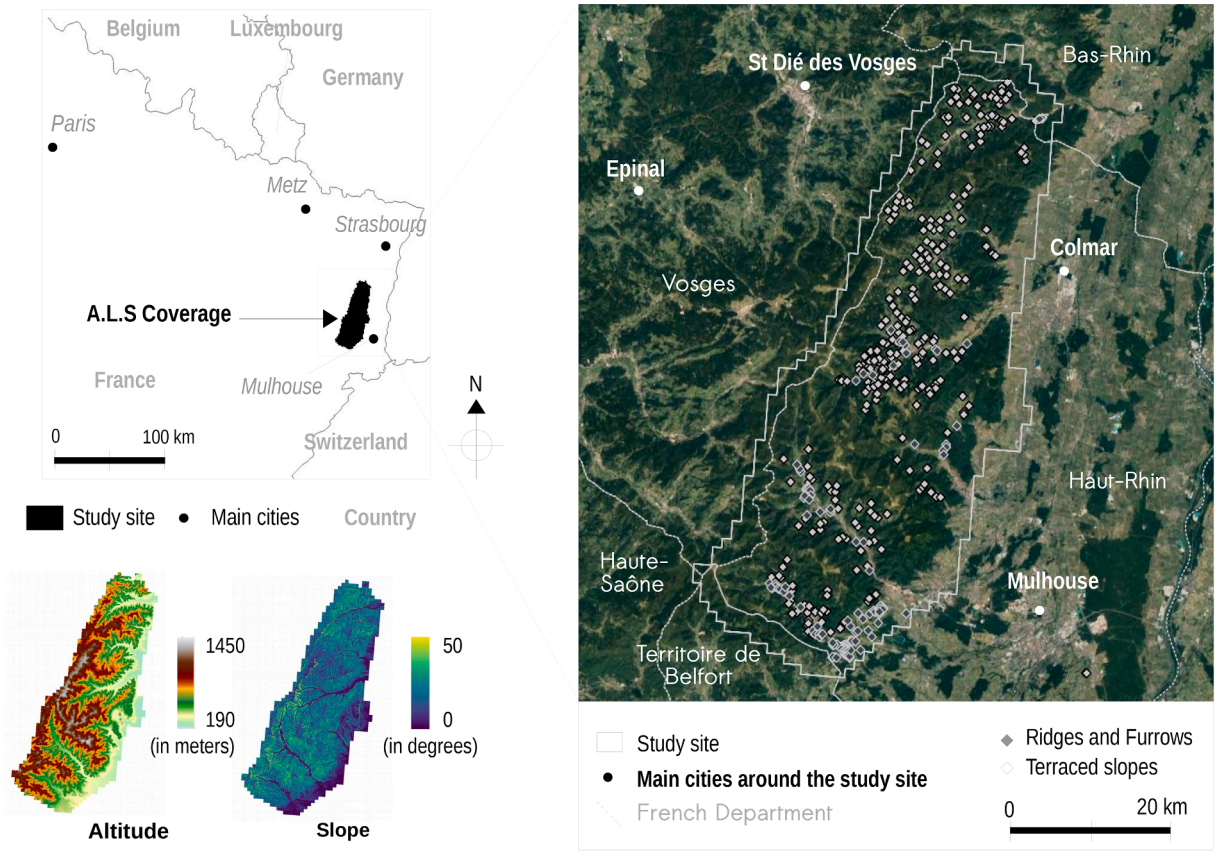


Fig. 1. Map of the study site, ALS coverage, and the location of known archaeological sites.

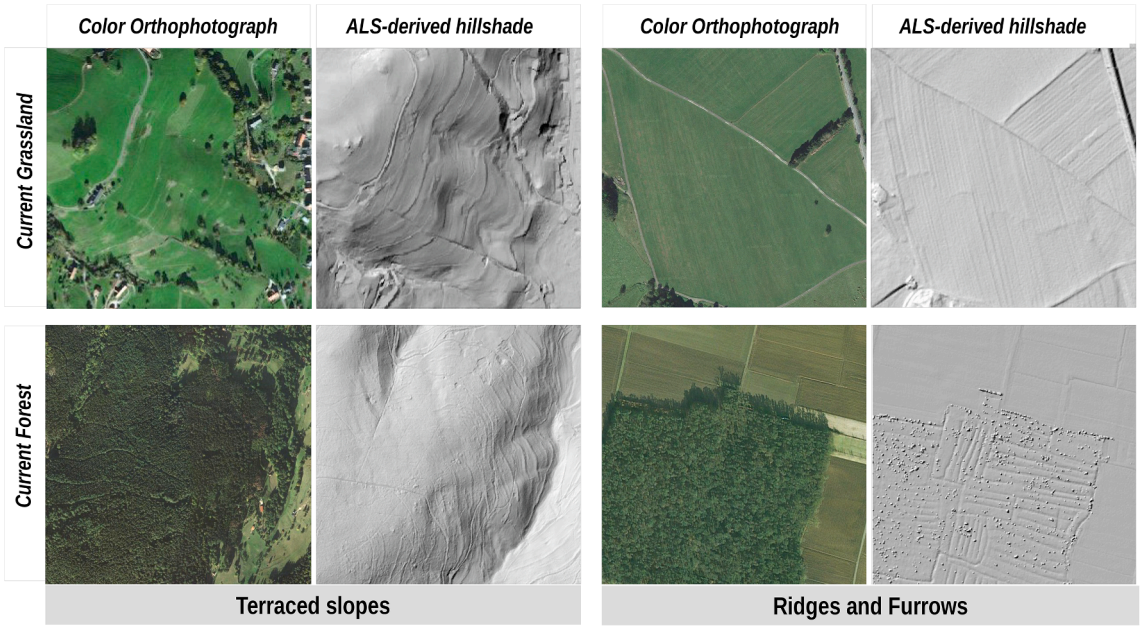


Fig. 2. ALS-derived hillshade examples of terraced slopes and ridges and furrows characterised by the present land cover, i.e., forests and grasslands.

produced by the French Mapping Agency (IGN, 2020).

From previous fieldwork and the visual inspection of the interpolated ALS data, we manually digitised every visible FFS on the entire study region, namely 305 TS and 112 RaF objects. Then, this dataset has been used as training and validation sets in the development of detection models, for the present study (see Section 2.5). The slopes of the TS varied from 6.7° to 30.9° (mean $= 18.6° \pm 4.7°$), and TS were found equally on southern and northern exposures (respectively 164 on south-facing and 142 on north-facing slopes). We also observed, however, a linear relationship between elevation and TS exposure; as TS elevation increased, the proportion of TS on north-facing slopes decreased (only 30% of these TS were found above 700 m). TS were covered mostly by grasslands (156/306 sites) and forests (130/306 sites). RaF were found mainly in the southern part of the study site and were markedly less common than TS (Fig. 1). They were observed mostly in relatively flat areas (average slope $= 4.7° \pm 2.8°$) at low altitudes (412 ± 47 m). Similar to TS, they were most commonly covered by grasslands (43/112 sites) and forest (28/112), but a high number of these features were covered by croplands.

2.4. Automatic Detection Processes

A global flowchart of the developed processing chain is presented in Fig. 3 and illustrates the three main steps of our approach: (1) pre-processing of the ALS-derived elevation data; (2) filtering and augmenting the learning set; and (3) training of the classification models

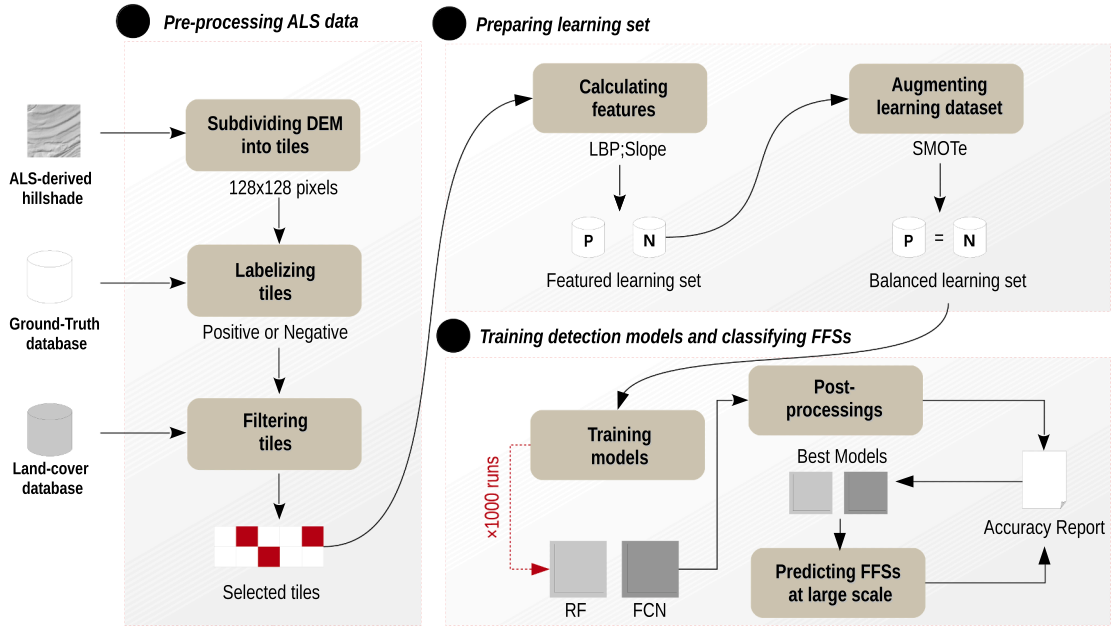


Fig. 3. Detailed flow chart of the methodology applied in this study to produce the automatic detection of former field systems (FFS) through machine learning and local binary patterns.

and predicting FFS.

2.4.1. Pre-Processing of the ALS-derived elevation data

Preprocessing involved initially subdividing the DEM into tiles of 128 × 128 pixels, respecting an overlap of 64 pixels between adjacent tiles. In total, 142 472 windows were generated. Size was defined empirically to cover a large subset of the surveyed forms to calculate features on images having marked patterns, i.e., both structuring (corners, lines) and background elements (fields). The overlap between windows served to include information occurring at the edges of images and to ensure continuity for the derived features. To label each tile, we intersected the centroid of each tile with the polygons of the identified sites. In total, for TS we found 6 565 positive tiles and 135 907 negative tiles, and for RaF we found 295 positive tiles and 142 176 negative tiles.

The second step eliminated "negative" candidates from the learning set to reduce the class imbalance caused by an under-representation of positive examples relative to negative examples. Background examples were also over-represented for multiple reasons (Trier et al., 2019). First, at the regional scale, the background is the most common situation in the landscape. Second, an archaeological object is, by nature, rare, thereby limiting the number of occurrences in a learning set. Therefore, the developed models tend to overdetect negative examples at the expense of the positive ones; the latter includes a significant part of the landscape characteristics. We performed three cleaning steps. (1) We used the OpenStreetMap landcover database to eliminate negative candidates crossed by linear objects, such as roads or pedestrian pathways. This ensured that we avoided confusing these objects with lines contained in the surveyed FFS, such as parcel boundaries. (2) We then used a regional landcover database covering the Grand-Est (CIGAL, https://data.geocatalogue.fr/id/dataset/FR-236700019-BdOCS20112012-CIGAL-V2) region to remove negative examples located on water or in artificial areas (buildings or roads). For the learning set, and based on our expertise, we assumed that it was impossible to observe the surveyed archaeological forms in these areas. (3) Finally, negative tiles located in the first-order neighbourhood of positive tiles were removed, as they likely shared similar features characteristic of positive tiles partially covered by FFS. After these steps, we had 44 612 TS (6565 positive and 38 047 negative) and 38 266 RaF tiles (295 positive and 37 971 negative) in each training set.

2.4.2. Preparing learning sets

To train the RF and FCN models, we calculated two categories of features from DEM tiles, namely LBP and slope metrics. LBP provide information regarding the spatial structure of a local image texture (Ojala et al., 2000; Ahonen et al., 2004; Li et al., 2015). The principle consists of labelling the neighbouring pixels g_P of one central pixel g_C and considering the result as a binary number (gives 0 if each pixel is smaller than the centre, otherwise 1). As the binary code of a LBP filter is retrieved with a fixed number of neighbours P, the resulting image is often represented with its histogram, which can be used as a texture descriptor (each histogram bin becoming one specific feature, Fig. 4). In the original set-up, this histogram comprises 2^p bins for a given radius r. LBP code is invariant to any monotonous grey-level transformation, and the local neighbourhood binary code remains unchanged after transformation. Hence, these properties make these filters very suitable for the analysis of textural images acquired under uncontrolled conditions, such as our ALS data. They also allow characterising structures located in variable topographic environments, such as high- or low/medium sloped areas.

The LBP code is expressed as the decimalised form of an octet binary number:

$$LBP_{P,R} = \sum_{p=0}^{n} s(g_P - g_C)2^p, \quad s(x) = \begin{cases} 1, x \geqslant 0 \\ 0, x < 0 \end{cases}$$

LBP were calculated from multidirectional hillshade layers. Eight directions were defined (0°, 20°, 45°, 65°, 90°, 115°, 135°, 170°), and we repeated the process for two illumination angles (15° and 315°). In total, 16 hillshade layers were calculated and merged into a band stack for each candidate. We then ran principal component analysis (PCA) to eliminate any redundant information. We selected the first two principal components to support LBP calculations (noted as u_1 and u_2). For each hillshade PCA band, 30 LBP features were calculated corresponding to 30 bins of the LBP histogram (noted as $[u_{1,2}\_0, \ldots, u_{1,2}\_29]$, Fig. 4).

In addition to determining the LBP variables, we calculated basic measurements for a slope (mean, standard deviation, maximum, and minimum). We assumed that sought-after structures were installed within specific topographical contexts. In particular, slope remained as a fundamental determinant for TS, as below or beyond a certain slope threshold, erosion can no longer shape lynchets. RaF may also exhibit

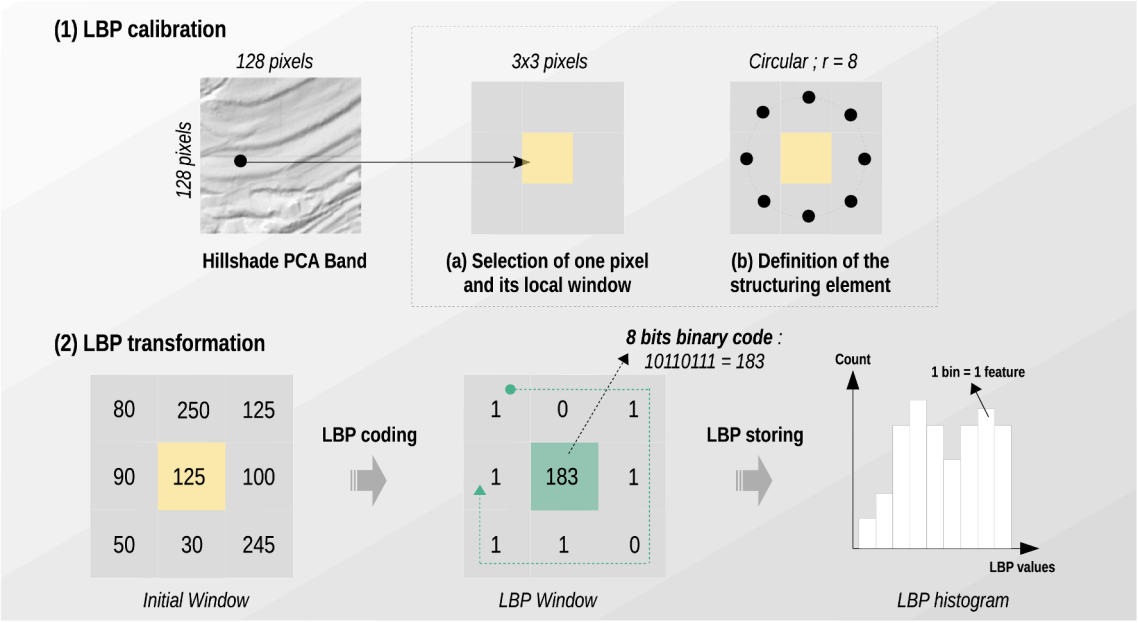


Fig. 4. Calculation of LBP features from the ALS-derived hillshade layers.

specific slope distributions because of alternating ridges and furrows.

To address the issue of unbalanced data sets, we used a synthetic minority oversampling technique (SMoTE; (Chawla et al., 2002; Douzas and Bacao, 2018)). This technique uses nearest-neighbour classification to create artificial observations of unbalanced classes that are similar to the existing ones. For archaeological purposes, the aim was to create new positive samples, normally under-represented in the feature space, until we obtained a uniform distribution ($N_{Positive} = N_{Negative}$). The main advantage of SMoTE, compared with traditional random naive over-sampling, is that by creating synthetic observations instead of reusing existing observations, overfitting of the classifier is reduced. We applied SMoTE to both data sets, i.e., TS and RaF; these data sets were then used to train and validate the classification models.

2.4.3. Training detection models and classifying FFS

We compared two classification models for detecting FFS. First, we trained a RF model due to numerous reasons. (1) The parameters are in a rather low number, and are also easy to understand. Thus, specialists such as archaeologists or historical ecologists can readily use the detection model. (2) RF offers the opportunity to measure the importance of input features variables. Under certain circumstances, this would help identify variables with low contribution, thus allowing training and prediction to run faster. (3) RF is able to manage missing data in case of an incomplete coverage for example. (4) RF offers a good comparative basis to deep learning models because of its ability to properly learn information with limited training datasets, as is often the case in an archaeological context (Breiman, 2001).

RF is a well-known ensemble learning method that combines kBinary with classification and regression trees (CART; (Breiman, 1984)). As an ensemble classifier, RF provides a classification score calculated from classification results of the individual and independent decision trees of the RF. In this study, hyperparameters, such as the number of decision trees in the forest and the number of features considered by each tree when splitting a node, were tuned by a random search method (Bergstra and Bengio, 2012). This algorithm runs all possible combinations of parameter values and returns the most accurate combination.

The random grid search allowed us to determine a first set of hyperparameters (Table 1). We then manually fine-tuned this set to obtain the optimal classification results. The final obtained hyperparameters were: (1) the number of estimators set to 8, (2) the maximum depth of the tree set at 24, (3) the minimum number of samples to split a node set at 2, and (4) the minimum number of samples to be a leaf node set at 2. The maximum number of features considered to find the best split was determined by the square root of the number of features.

Second, we trained a FCN to assess the potential of DL methods for the detection of FFS. Unlike CNNs, FCN consists of a series of fully connected layers, but does not contain any convolutional layers making its architecture more simple. Moreover, as we did not use transfer learning for the reasons previously described in Section 1, we hypothesised that FCN was likely to learn a bigger amount of information from the training set before overfitting. This would more likely not be the case with CNNs, due to the limited dataset. FCN also offers the opportunity to use pre-computed features such as LBP. Therefore, we made the assumption that combining LBP features and the SMoTE technique (see Section 2.4.2) was an effective solution for increasing the training set and tackling issues caused by the unbalanced nature of the data.

For the FCN, we first optimised the hyperparameters using a random search strategy. We searched only for the number of hidden layers and neurons (Table 2).

In addition to inputs and outputs, the final model consisted of four hidden layers, containing 20, 15, 10, and 5 neurons, respectively. A rectified linear unit (ReLU) was used as the activation function for each layer. To improve model performance during the learning process, we used batch normalisation to normalise the output of each layer. A common technique consists of using dropout layers to prevent overfitting by the model. By removing a certain number of neurons, the network learns more robust parameters and generalises better on yet-unseen data sets. However, considering the large imbalance between positive and negative observations for our labelled images, the dropout approach was deemed inappropriate, as it could result in a loss of crucial original (non-augmented) data. Finally, the final layer returns the probability of the image belonging to one of the two classes (positive or negative).As we were dealing with a binary classification problem, we applied a binary cross-entropy loss function to compute the prediction error of the model. In the context of this study, the loss value allows comparing the image label with both probabilities that the model returns as an output. The learning rate, which is the rate at which a neural network updates its parameters, was determined automatically for each data set (RaF and TS) using the FastAI library (Howard, 2018).

After defining the hyperparameters, we used the FastAI library to evaluate the evolution of the difference between the validation and the training losses for both models (i.e. RaFs and TS). Preliminary tests showed that models would overfit past 30 epochs, and underfit before reaching 10 epochs. Therefore, an early-stopping technique was used (Prechelt, 1998). Moreover, the number of epochs was later fixed to 25 for a better comparison between models obtained at different steps of the bootstrap technique used for accuracy assessment. This process is described in Section 2.5. Also, this number appeared to be a good compromise to maximize the amount of information learnt by models, while avoiding overfitting issues.

2.4.4. Post-Processing

Classification outputs were filtered following a spatial continuity criterion. As TS and RaF sites spanned over a minimum of two adjacent tiles, we thus assumed that an isolated tile detected as “positive” by the classifiers corresponded to an erroneous detection. From this perspective, a first-order rook contiguity was applied to the classification outputs. In other words, each tile detected as “positive” was inspected iteratively and then accepted or rejected for this spatial criterion.

2.5. Accuracy assessment

We used a repeated holdout cross-validation method to assess classifications to evaluate the performance of the classification models on similar data sets. We used only non-synthesised candidates for this evaluation. Thus, for each run (100 in total), 80% of the training set was randomly selected and used to train the models; 20% of the training set was used for predictions. We determined the average F-score and explored Recall and Precision scores to provide more insight into the classification results (Sammut and Webb, 2011).

Table 1
Random Forest’s hyperparameters values used in this study.

Hyperparameter	Minimum value	Maximum value	Steps
Number of estimators	1	200	10
Maximum depth of the tree	None (infinite depth), 10	60	10
Minimum number of samples to split a node	2	10	2
Minimum number of samples to be a leaf node	1	10	2
Number of features considered for a split	log(N), sqrt(N)		

Table 2
FCN’s hyperparameters values used in this study.

Hyperparameter	Minimum value	Maximum value	Steps
Number of hidden layers	1	8	1
Number of neurons for each layer	5	50	5

In a second step, we evaluated detection results over the entire study site. Hence, we applied the best-performing RF and FCN models to the full data set on the basis of the obtained F-score after 100 iterations. From a semantic point of view, true- and false-positive identifications rates (TPI and FPI) were used to compare our results with the existing literature. To spatially evaluate detection results, we reported the sum of positive predicted tiles in a 1 km × 1 km grid and compared these results with the reference. We also calculated a structural similarity (SSIM) index between the automatically derived and reference spatial patterns to measure the capabilities of our approach for retrieving the major spatial trends in the landscape.

Finally, we assessed the scale of the FFS sites, i.e., for a single digitised TS or RaF polygon. This evaluation aimed to determine whether the developed processing chain could detect the presence of FFS instead of necessarily retrieving their entire spatial extent. We strongly believe this evaluation is crucial as the developed method can serve as a FFS detector by archaeologists or landscape ecologists, rather than considering it as an accurate spatial delimiter of FFS. In this respect, a FFS object was deemed as detected when more than one detected positive tile was contained in the latter object.

3. Results

3.1. Cross-validation derived results

Fig. 5 illustrates the average classification scores obtained after 100 iterations according to random validation sets (split 80–20%). Classification after post-processing for TS showed a mean F-score of 0.84 ± 0.09 for both models. For the RaF data set, average results exhibited poorer performance than for TS, producing a mean F-score of 0.58 ± 0.08. We noted a greater variability in the classification scores after the 100th iteration in the RaF configuration (mean SD F-score$_{RaF}$ = 0.0555 ± 0.2 and mean SD F-score$_{TS}$ = 0.0075 ± 0.2). These first results were expected given the low fraction of positive RaF cases (295 positive tiles among 142 472 tiles or 0.20%) when compared with the fraction of positive TS examples (6565 positive tiles among 142 472 tiles or 4%).

RF performed, on average, better than FCN for both data sets after post-processing. In the TS configuration, RF had an average F-score of 0.91 ± 0.006, whereas the average F-score of the FCN model was 0.77 ± 0.006. Similarly, RF produced an average F-score that was 24% higher than that obtained via the FCN model for the RaF configuration (0.64 ± 0.058 and 0.52 ± 0.053 for RF and FCN, respectively). Nonetheless, average recall and precision scores showed more balanced results, mainly for the RaF data set. Although the average precision score confirmed the superiority of RF over FCN for the RaF data (average Precision$_{RF\text{-}RaF}$ = 0.93 ± 0.051 and, average Precision$_{FCN\text{-}RaF}$ = 0.47 ± 0.069), average recall scores exhibited the reverse trend. The ability of FCN to correctly detect positive RaF tiles was greater than RF (Recall$_{RF\text{-}RaF}$ = 0.49 ± 0.064, and Recall$_{FCN\text{-}RaF}$ = 0.59 ± 0.060). In the TS configuration, average recall and precision scores remained higher for RF and confirmed the superiority of the latter model over the FCN model.

We observed significant improvements to the classification results when we applied post-processing. For both data sets (i.e. TS and RaF) and regardless of the selected model, F-scores improved by 66% on average after post-processing. These results, however, must be nuanced in the context of the other classification scores. Although average precision scores were improved for TS and RaF in both classification models, average recall scores were negatively affected. Thus, post-processing limited the overdetection of negative examples; at the same time, it decreased the rate of correctly detecting positive tiles. This was notably the case for the RaF data set where we observed a recall score decrease of approximately 0.41 points for both models. It should also be noted that the average recall score obtained with FCN for RaF detection without post-processing was equal to 1. All positive validation tiles were classified correctly at the expense of a very low precision score.

We also studied the variable importance ranking, measured by the mean decrease in accuracy (MDA) for the 30 most important variables of the RF model (Fig. 6). The most important variable for classifying TS was mean slope followed by LBP bins, such as u_1_29 and uu_1_1. Similar metrics from the second hillshade component were also ranked sixth and seventh (i.e., u_2_29 and u_2_2). The latter features correspond to the edge-like regions of TS, indicating that LBP are particularly sensitive to repetitive-oriented parcel boundaries. Metrics that represent primarily flat-like regions were ranked fourth and ninth, which demonstrates that regular patterns between hedges also influenced the resulting LBP metrics. LBP bins corresponding to flat-like regions were also the most represented in the top ten variables for RaF, with u_1_16 being the most important variable used by the model. u_1_13 and u_1_15 were ranked fourth and fifth, respectively. Slope features also played a significant role, as mean slope (*mean_s*) was ranked as the second most important variable, and SD slope (*sd_s*) was seventh.

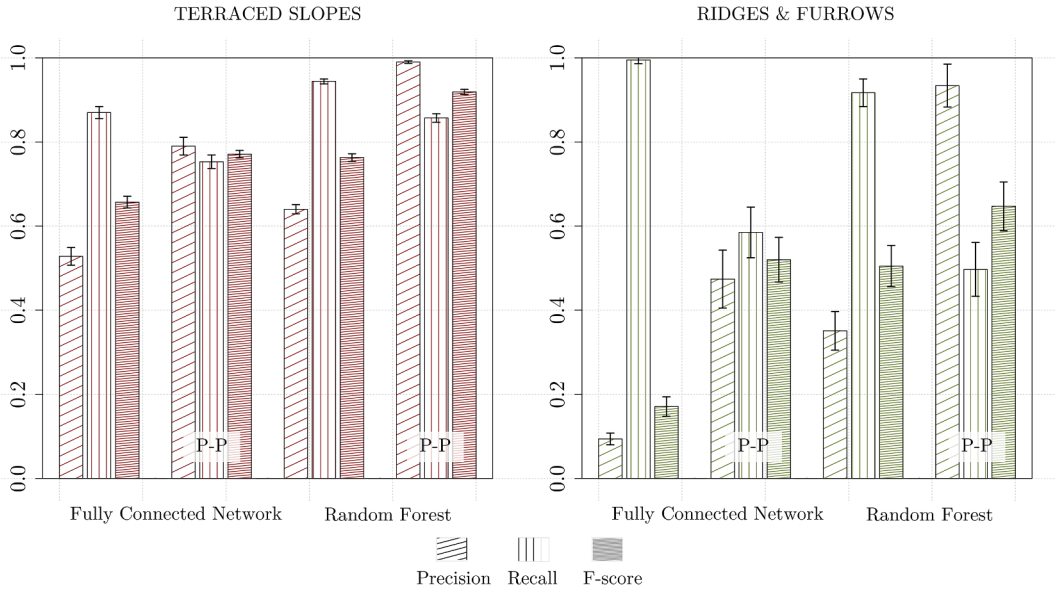


Fig. 5. Bar plots illustrating the mean classification scores for the FCN and random forest (RF) models after the cross-validation stage. Error bars represent the standard deviation calculated from 100 iterations. "P-P" indicates that post-processing was performed after the classification step.

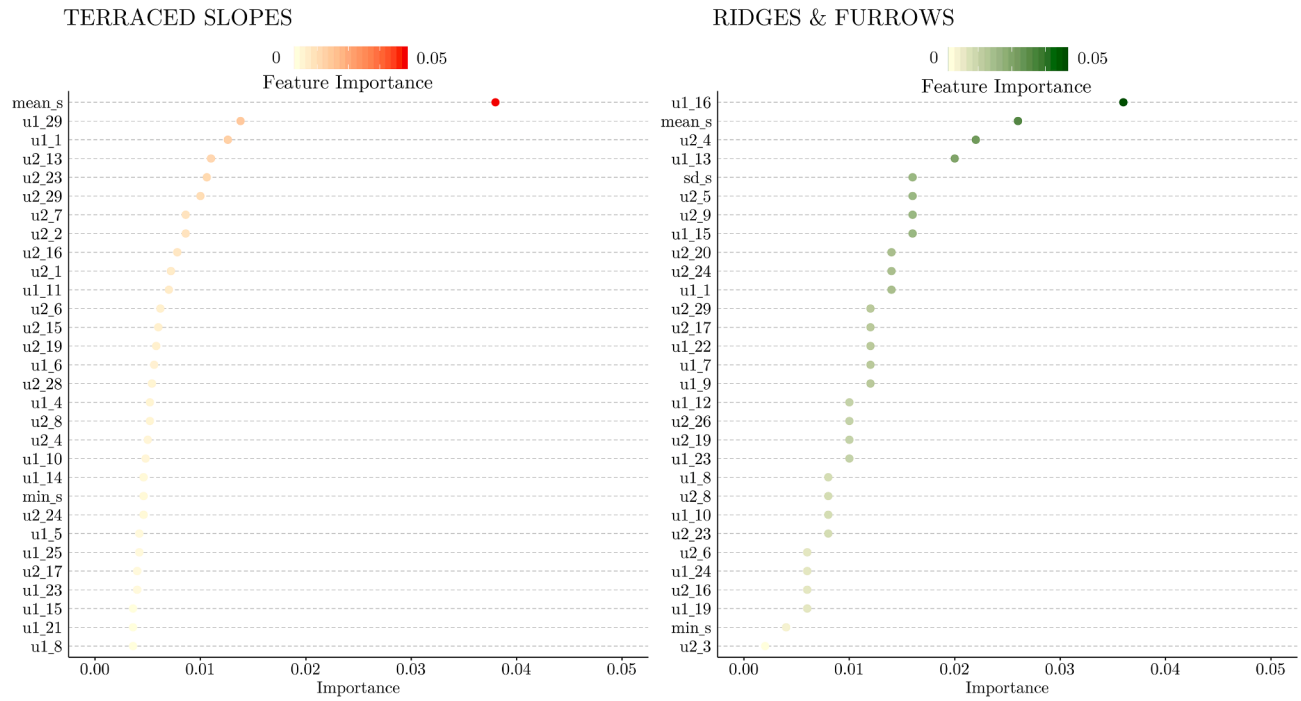


Fig. 6. Importance ranking of variable on the basis of the mean decrease in accuracy (MDA) scores using the random forest model.

3.2. Prediction results over the entire study site

Large-scale results for the TS and RaF data sets are shown in Table 3. We focus here on the results for detecting positive tiles to examine the ability of the developed processing chain to retrieve objects of interest over large territories. Furthermore, we also focus on the positive tiles to increase the transparency of our results and facilitate comparison with similar previous or future findings (Fig. 7). The classification results discussed here were obtained using the best-performing RF and FCN models selected after the cross-validation stage.

Classification results showed that 4973 of 6565 TS tiles (75%) were identified by FCN with 1289 (20%) false positives. We observed, however, a more robust performance for the RF model with 5880 tiles (89%) assigned to the TS class. These TS predictions are quite meaningful because many true locations were identified, and the number of false positives was relatively low (maximum = 20%). The results also confirmed previous results derived from cross-validation, demonstrating the high capabilities of the developed approach for locating TS at a large scale.

RaF predictions appeared more chaotic. Only about half of the 296 RaF tiles in this block were correctly classified. We were able to observe that FCN provided slightly better results than RF, producing a FPI rate of 57% versus 51% for RF. However, the number of false predictions varied greatly between models. Although the FPI rate remained high for the FCN (60%), we observed satisfactory results for RF with a FPI rate equal to 2%.

Table 3
Classification results of random forest and FCN models for identifying terraced slopes (TS) and ridges and furrows (RaF) of former field systems (FFS) across the entire study region.

FFS	Model	True positives	False positives	Known FFS tiles
Terraced slopes	RF	5880 (89%)	15 (0.2%)	6565
	FCN	4973 (75%)	1289 (20%)	
Ridges and Furrows	RF	151 (51%)	7 (2%)	295
	FCN	169(57%)	180 (60%)	

Maps of predicted tiles per 1 km × 1 km are shown in Fig. 8. Both RF and FCN models provided very similar spatial patterns to the references for the TS data set ($SSIM_{RF/Reference} = 0.98$, $SSIM_{FCN/Reference} = 0.95$). Similarities between the reference spatial patterns and the model predictions were lower for the RaF dataset but were nonetheless satisfactory ($SSIM_{RF/Reference} = 0.83$, $SSIM_{FCN/Reference} = 0.75$).

These statistical maps also showed that FCN overdetected tiles marked by a higher density of squares containing a low number of tiles (between 1 and 5) than in the reference map (see histograms). Although these errors are of note, it is also important to mention that FCN exhibited a better generalisation power than RF. This was particularly visible for TS in the south-eastern area of the study site, where slope gradients are low and differ from areas that contain most known sites. For these latter areas, FCN correctly detected more tiles than RF, indicating a greater ability to predict TS within variable topographic conditions, less represented in the learning set. For RaF, both models detected tiles corresponding to major high-density zones observed in the reference layer. However, overdetected tiles were particularly numerous with FCN in the extreme eastern part of the study site. After visual inspection, these errors were instructive because these areas corresponded to actual RaF sites that were undetected during the ground-truthing stage. Thus, this result may highlight the strong ability of FCN to survey unseen data and its low sensitivity to annotation errors.

Detection results reported at the scale of the FFS sites showed similar results to the classification results derived from the window scale. From a total of 305 TS sites, 249 (81%) were found by FCN, whereas 261 (85%) were detected by RF. However, less than half the known RaF sites were located by FCN (45 sites of 112 (40%) and only 39 (35%) sites via RF. Thus, errors were grouped mainly within specific zones that were, in general, poorly identified by the classification models. These zones corresponded to undetected full sites and a few isolated sites found within the landscape or located at the margin of identified sites.

4. Discussion

Overall, our results demonstrated the high suitability of our approach for detecting terraced slopes and ridges and furrows across the landscape. The combination of LBP texture operators and ML techniques

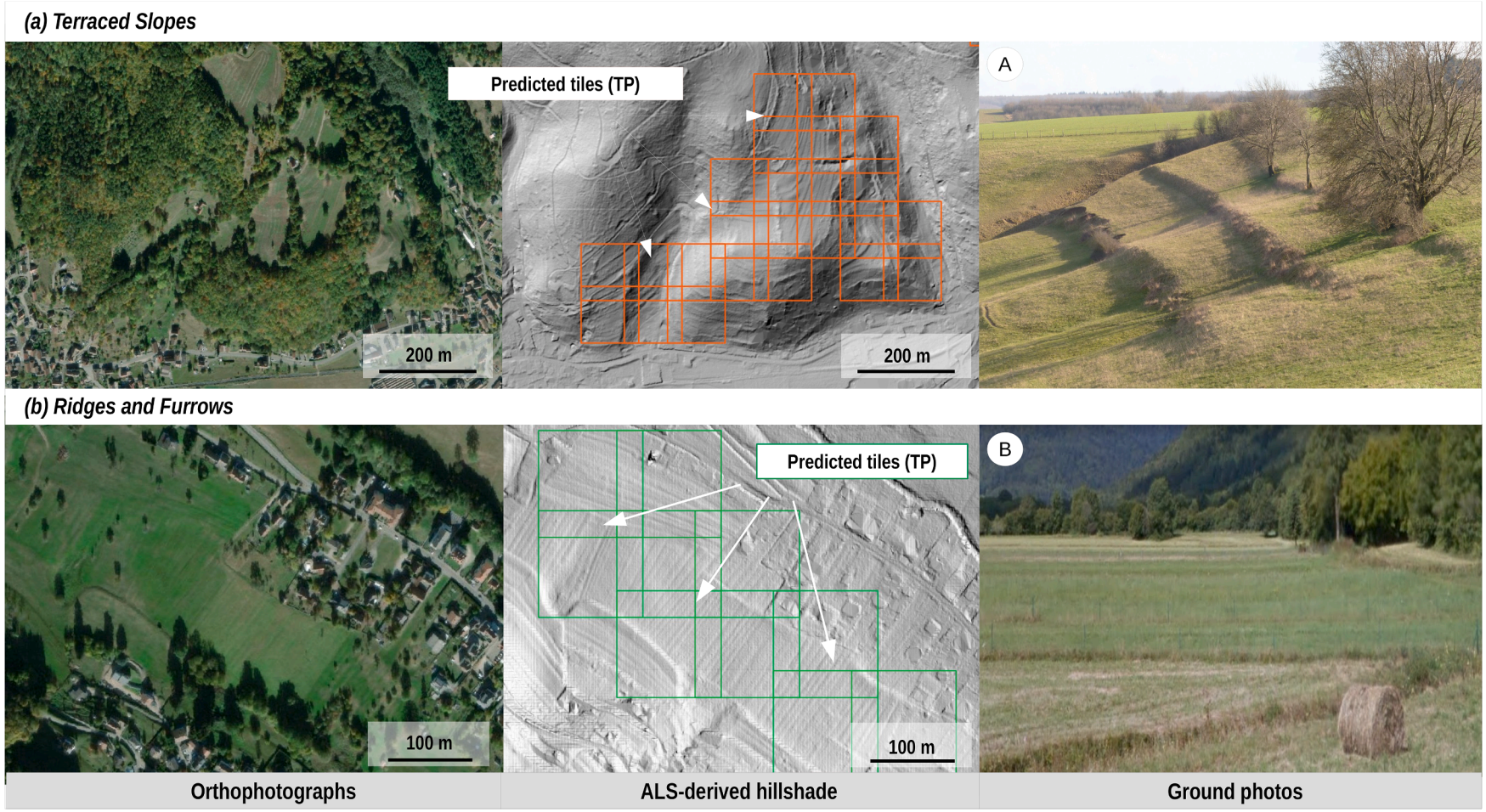


Fig. 7. Examples of detected tiles (TP = True Positives). predicted by the best FCN models and corresponding archaeological structures. Results are shown for the municipalities of Breitenbach (A; 7°09'41.82"E–48°02'59.85"N) and Luttenbach (B; 7°12'24.71"E–48°04'38.83"N).

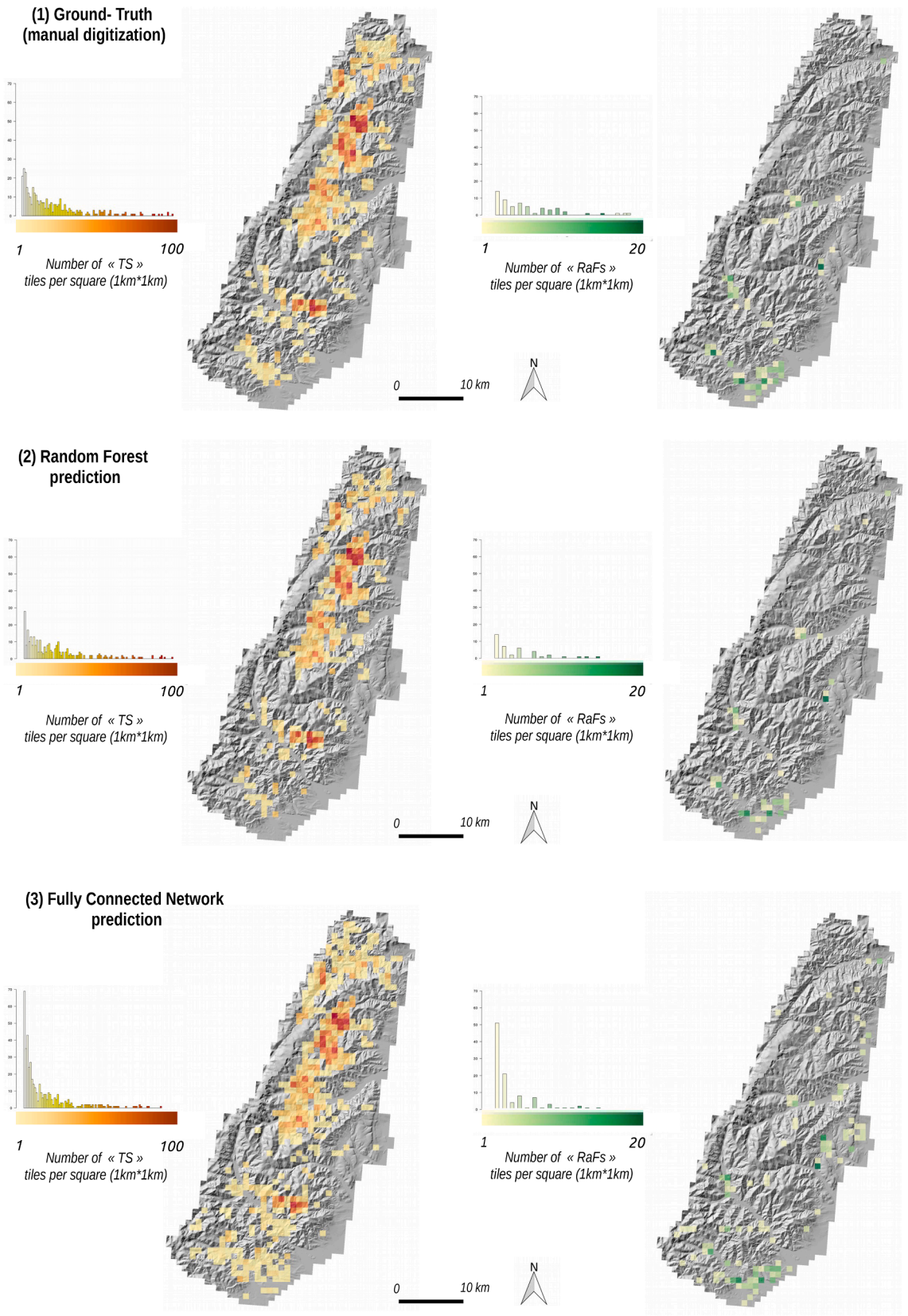


Fig. 8. Prediction maps of TS and RaF over the entire study site for the different detection models against validation maps. Predicted positive tiles (128 m × 128 m) were summed for each 1 km × 1 km cell. Colour gradients illustrate positive tile variations over cells, whereas histograms show cell densities according to their number of positive tiles.

offers a promising framework for mapping ancient agricultural landscapes from ALS data. To the best of our knowledge, our study is the first to use LBP and train a FCN model to detect archaeological objects on the basis of ALS data.

In particular, our approach reproduced larger trends in the landscape with a high level of similarity between the predicted and reference spatial patterns of FFS. These results are in line with the recommendations of (Cowley, 2012), who emphasised the importance of providing large-scale maps for identifying the key structures of surveyed forms with reasonable areas of uncertainty due to the complexity of the investigated landscapes. Freely distributed ALS data, such as RGE-Alti, are sufficiently adapted to the needs of detecting archaeological sites. These data were initially acquired for different purposes than archaeological ones, such as forestry or water monitoring, and their quality could be questioned, namely due to moderate pulse densities and/or insufficient spatial resolution. Our study demonstrates that these potential issues do not prevent their use in detecting FFS but also that large-scale RGE-Alti missions provide crucial information for investigating the spatial patterns of FFS at a large scale. Moreover, because of the broad area of France covered by these data, this approach could be run on any RGE tile suspected of containing historical agricultural activities. This study therefore provides a novel avenue for mapping ancient landscapes at a national–international scale. To go even further, we also believe that different spatial outputs should be proposed by exploring the inherent properties of the classification models, such as probability scores in the decision process. Fig. 9 illustrates an example of a probability map derived from the FCN model, without any post-processing, for identifying TS. Probability gradients allow for the rapid identification of tiles of interest, with a high level of certainty, to direct specific field investigations, such as archaeological dating as well as topographical and ecological surveys. This approach can also provide more nuanced maps for landscape archaeologists and ecologists, then combining them with external data (excavation and soil data) to help to confirm the presence of these landforms.

Regarding the two surveyed FFS, cross-validation results initially demonstrated better performances in the detection of TS compared with RaF. Despite a larger intraclass heterogeneity among TS, these results were expected because of the low number of known positive examples in the RaF training set. The small amount of training data is a well-identified issue in archaeological site detection (Trier et al., 2019; Verschoof-van and Lambers, 2019). Moreover, background tiles are, by nature, over-represented compared with archaeological objects when processing large areas; this over-representation reinforces the strong imbalance in the learning set, which makes this task challenging and more complicated than general image classification. To address this issue, we proposed two steps in this study that should be considered as prerequisites in future research of a similar nature. First, tiles should be filtered according to assumptions related to the landscape. This step allowed an effective reduction in the number of background examples. Second, the use of data augmentation, such as via the SMoTE technique, can rebalance training sets and improve classification. The latter method has the advantage of requiring few parameters and being independent of the classification technique. We also strongly believe that efforts to increase archaeological datasets should be pursued in three lines of works: (1) to keep on developing citizen web platforms in order to increase labelization of archaeological forms (see https://www.globalxplorer.org/); (2) to keep acquiring ALS data on vast territories in order to conduct large-scale non-intrusive archaeological surveys and to obtain uniform image datasets; (3) to keep developing GIS approaches to align existing ALS-derived data (co-registration, disaggregation, etc.) and to maximise the amount of available sources.

In addition to these steps, several precautions also be taken. In the RaF configuration specifically, the SMoTE technique might also explain a part of classification errors. Because of a very low number of positive samples and the random sampling, the data space in which additional positive samples are calculated may be underestimated. Consequently, new samples increase within a very restricted data range that does not cover the full range of positive samples. We thus suggest evaluating the use of SMoTE in archaeological surveys regardless of the selected classification model (i.e., early ML or DL models). Nevertheless, the collaboration between landscape archaeologists, ecologists, and remote sensing specialists is required to increase the number of known positive sites across different geographic regions and to fully capture all possible configurations. Another solution is to increase the number of examples by calculating features of similar actual landscape forms (Cowley et al., 2019). For instance, it may be useful to calculate LBP metrics on aerial greyscale photographs, including parcels networks, to complete the RaF training set.

For the classifiers, RF outperformed FCN with a higher F-score obtained for the TS data set. This was confirmed by the accuracy assessment reported for the entire study area. Particularly, the false-positive rate with RF was substantially lower than that of FCN. These results corroborate previous findings, which showed strong RF performance when the training data are well represented (Menze and Ur, 2012; Guyot et al., 2018); however, these results should be nuanced. First, we applied an 80%–20% split to fully capture trends in the training set. Therefore,

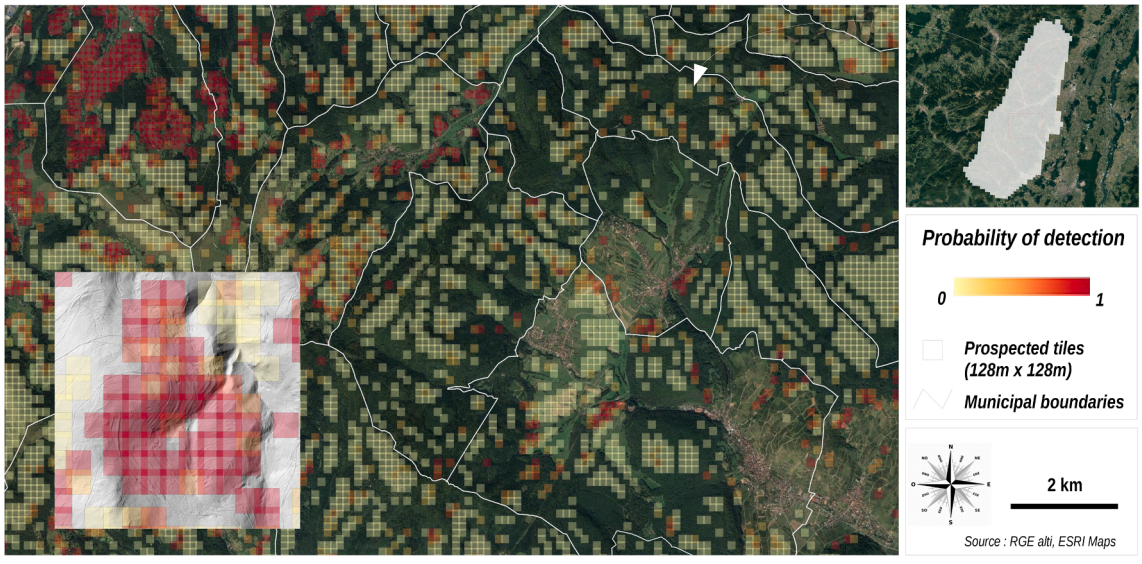


Fig. 9. Example of TS probability map generated through the properties of the FCN model decision process.

classification results presented for the full data set are only based on 20% independent tiles. This could overestimate accuracy indicators, as the latter model is known to fit the training data very well. Second, as the mean slope was the most important predictor for this model, this model is consequently highly dependent on landscape topography, which varies greatly among geographic regions. Therefore, transfer learning and/or domain adaptation are essential and should be further tested to extrapolate the trained model to unseen data.

RaF detection results showed that FCN was able to better locate positive RaF tiles than RF but with a moderate precision score. There may be several reasons for this pattern. The first is that most false negative errors were located in flat areas (eastern part of the study area) or in regions where the DEM contained some triangulation errors resulting in crossed straight lines that appear similar to RaF parcel boundaries. These triangulation errors currently occur for south-facing slopes covered by actual forest, which may limit pulse densities during the acquisition stage because of a higher density of deciduous trees (Trier and Pilø, 2012; Pregesbauer, 2013). Adapted methods could therefore be used to prevent interpolation errors in the DEM and tackle irregular point densities across a region of interest. Another option would be to introduce exposure and actual land cover into the model as new features to provide new pertinent information to the FCN decision process. However, overfitting issues may appear because of too many features relative to the small number of learning samples, a scenario that is common to archaeology. The second factor is that a large portion of these errors corresponds to actual RaF sites that were not detected during the ground-truthing stage. These findings are particularly meaningful as they show that FCN can adapt appropriately to previously unseen data despite a low number of training samples. They also show that FCN is minimally influenced by the remaining labelling errors, a recurring issue when ground-truthing is performed manually, i.e., manual digitisation. These latter observations are promising and show that FCN can provide a credible alternative in similar future studies.

We also highlighted the difficulties of performing an effective and generic post-processing stage for multiple FFS. Although the recall decrease remained acceptable, all the while maintaining a high precision in the TS configuration, this led to more chaotic results for RaF. Two main reasons can be given: (1) the minimum number of connections (three) in the rook contiguity for classifying a tile as positive is too restrictive in relation to the limited extents of RaF sites. Adapting this parameter according to the surveyed FFS, for instance $n = 1$ for RaF, would avoid converting tiles from positive to negative mainly when located in the borders of small isolated RaF sites; (2) the iterative process does not allow the simultaneous consideration of all combinations. Therefore, a tile may be converted to negative if it lies adjacent to a single negative tile and is processed initially, even if it forms part of a three-tile bundle. To face this limitation, one can perform several permutations in the iterative procedure to capture the different combinations possible in a tile bundle. Hence, each tile would be considered several times, and its belonging to a tile bundle would be better accounted for into the decision process.

In terms of the features, LBP were successfully used to discriminate FFS from background tiles, which offers an appealing alternative or complement to standard visualisation techniques (VTs). Indeed, the latter techniques are strongly dependent on empirical parameters, namely orientation, and thereby bias the analyses of archaeological remains of specific sizes or anomalies. In contrast, only a structuring element is needed for calculating LBP and inherent properties, such as invariance to rotation, and monotonic illuminations make these features generic and robust for any ALS-derived DEM. We also believe that the recent proposals of (Guyot et al., 2018) related to multiscale topographic indices are promising because they maximise both local and neighbouring information around surveyed objects. From this perspective, LBP might be adapted further to a multiscale issue to integrate landscape information at a wider scale for the description of FFS (Chan et al., 2007; Lin and Qi, 2015). Therefore, the relative importance of features in the decision process of RF shows that LBP features are considered as important predictors. The intermediate and final bins of the LBP histograms, i.e., u13–15 and u29, were particularly discriminant.Fig. 10 highlights flat, edge-like, and corner-like regions for a positive and negative TS tile. Histograms displaying the distribution of the pixels for a TS tile show higher pixel densities in the central (edge of linear objects) and the extreme regions (areas between line objects) than histograms corresponding to a non-TS tile. Therefore, LBP metrics are highly influenced by oriented lines created by the top of scree slopes and/or regions in between. With this in mind, LBP might be applied to any similar surveyed archaeological field systems, such as clear cairns or Celtic fields. Other archaeological remains characterised by punctual geometric shapes that are visible in ALS data, such as charcoal burning platforms or burial mounds, could also benefit from this approach, and LBP for these purposes should be explored further.

This study demonstrated the possibility of mapping ancient landscapes that date from the medieval/post-medieval periods at a regional scale. The potential insights for historical ecology studies are multiple. Historical maps are commonly used to provide baseline information prior to remote sensing when reconstructing landscapes and measuring the effect of changes on existing biodiversity. This strategy has been used successfully to demonstrate the historical effects of grasslands and forests on the current distributions of plant (Vellend et al., 2006; Cousins and Eriksson, 2001; Cousins et al., 2007), insect (Bommarco et al., 2014; Herrault et al., 2016), and bird species (Szabo et al., 2011; Ford et al., 2009). Until now, studies exploring the impact of more ancient land use/land cover, prior to the first maps, on present ecological patterns relied mainly on chronosequences (Fuller et al., 1998; Foster et al., 2003; Powers, 2004; Niklasson and Granström, 2000), opportunistic historical datasets (Bommarco et al., 2014; Jamin et al., 2020), and local ALS data (Dupouey et al., 2002; Georges-Leroy and Montes, 2013). Consequently, conclusions are often spatially limited.

The present study addresses these limitations by offering the possibility to assess a new spatiotemporal baseline and detect potential time lags over a longer period at a regional level (Fig. 11). More specifically, several ecosystem studies might benefit from this approach. For instance, the potential effects of former land use on forest plant communities could be explored at a regional level. Given those plant communities that develop after afforestation of abandoned lands are likely to differ, it is interesting to evaluate similarities between plant composition in forests masking FFS (or not). Our approach would allow identifying several groups of forests characterised by different histories (Fig. 11), whereas forest plant composition (taxonomy and functionality) might be assessed by exploring a free, large-scale botanical data set (Vigie Flore, TeleBotanica, TRY database). At this scale, the low granularity of free botanical surveys would not be an obstacle, and model outcomes would offer the possibility of verifying the existence of time lags in forest plant composition at an unprecedented spatial scale.

5. Conclusion

We developed a processing chain to detect Medieval Terraced Slopes and Ridges and Furrows in the Vosges mountains using an ALS-derived DEM. Our results demonstrated the high suitability of our approach for detecting FFS at a large scale and for reproducing major trends in the landscape. The proposed development relied on three specific contributions: (1) the calculation of LBP helped describe FFS efficiently, thus providing proper features for the training of detection models; (2) LBP features associated with the SMoTE technique proved to be a powerful combination for processing unbalanced and small limited archaeological datasets; (3) FCN are promising detection models for non-intrusive archaeological surveys at a large scale. Their simple architecture offers a certain flexibility and they showed a low sensitivity to annotation errors. We suggest that this approach has the potential to offer new spatiotemporal outlooks in Historical Ecology studies as a means of providing a new temporal baseline to study the effects of landscape

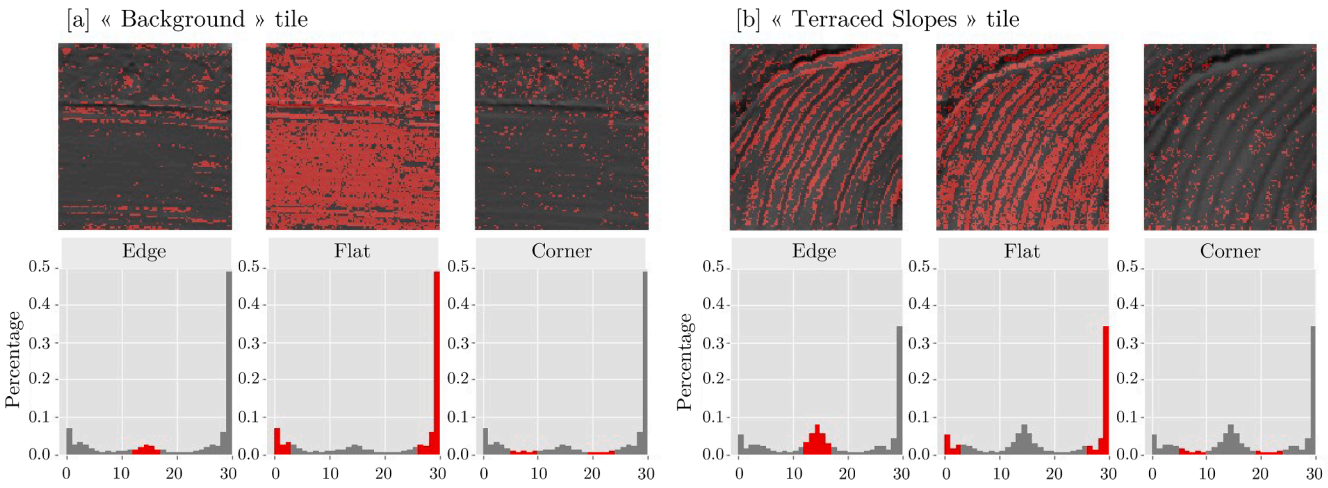


Fig. 10. Histograms of local binary patterns (LBP) highlighting pixel densities in flat, edge-like, and corner-like regions for a terraced slope (TS) and a background tile.

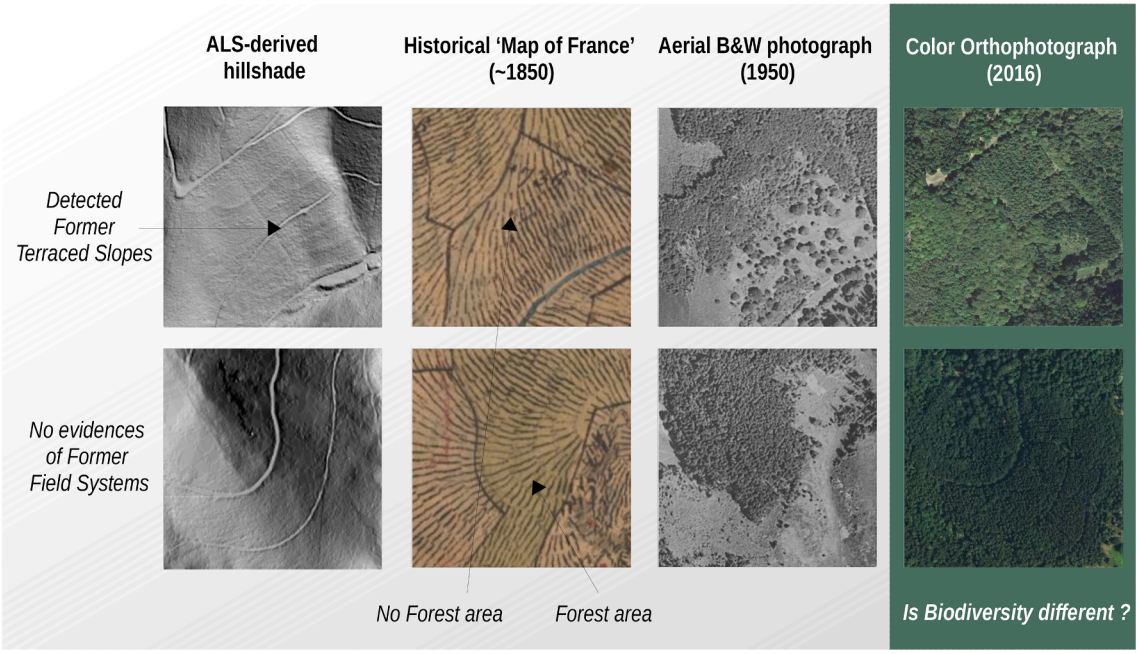


Fig. 11. Forest continuity trajectories based on a LiDAR spatiotemporal baseline. At the top, the reconstructed sequence highlights young forest areas with open environments, ca. 1850, and terraced slopes visible on ALS data. These patterns might indicate a significant presence of plants having open area–specific traits. Along the bottom, the forest area is quite visible in 1850, with no evidence of previous agricultural activities at this site. The forest continuity is high; therefore, forest specialist species should be dominant.

history on the current biodiversity patterns.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was funded by an IDEX-Attractivite fellowship (Excellence Initiative) provided by the University of Strasbourg (France). ALS data were obtained from the French National Mapping Institute (IGN).

References

Adriaens, D., Honnay, O., Hermy, M., 2006. No evidence of a plant extinction debt in highly fragmented calcareous grasslands in Belgium. Biological conservation 133 (2), 212–224.

T. Ahonen, A. Hadid, M. Pietikäinen, Face recognition with local binary patterns, in: European conference on computer vision, Springer, 2004, pp. 469–481.

Albrecht, C.M., Fisher, C., Freitag, M., Hamann, H.F., Pankanti, S., Pezzutti, F., Rossi, F., 2019. Learning and recognizing archeological features from lidar data. In: 2019 IEEE International Conference on Big Data (Big Data). IEEE, pp. 5630–5636.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. The Journal of Machine Learning Research 13 (1), 281–305.

Bommarco, R., Lindborg, R., Marini, L., Öckinger, E., 2014. Extinction debt for plants and flower-visiting insects in landscapes with contrasting land use history. Divers. Distrib. 20 (5), 591–599.

L. Breiman, Algorithm cart, Classification and Regression Trees. California Wadsworth International Group, Belmont, California.

Breiman, L., 2001. Random forests. Machine learning 45 (1), 5–32.

Bürgi, M., Li, L., Kizos, T., 2015. Exploring links between culture and biodiversity: studying land use intensity from the plot to the landscape level. Biodiversity and conservation 24 (13), 3285–3303.
Caspersen, J.P., Pacala, S.W., Jenkins, J.C., Hurtt, G.C., Moorcroft, P.R., Birdsey, R.A., 2000. Contributions of land-use history to carbon accumulation in us forests. Science 290 (5494), 1148–1151.
Chan, C.-H., Kittler, J., Messer, K., 2007. Multi-scale local binary pattern histograms for face recognition. In: International conference on biometrics. Springer, pp. 809–818.
Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research 16, 321–357.
Compton, J.E., Boone, R.D., 2000. Long-term impacts of agriculture on soil carbon and nitrogen in new england forests. Ecology 81 (8), 2314–2330.
Cousins, S.A., 2009. Landscape history and soil properties affect grassland decline and plant species richness in rural landscapes. Biol. Conserv. 142 (11), 2752–2758.
Cousins, S.A., Eriksson, O., 2001. Plant species occurrences in a rural hemiboreal landscape: effects of remnant habitats, site history, topography and soil. Ecography 24 (4), 461–469.
Cousins, S.A., Ohlson, H., Eriksson, O., 2007. Effects of historical and present fragmentation on plant species diversity in semi-natural grasslands in swedish rural landscapes. Landscape ecology 22 (5), 723–730.
D.C. Cowley, In with the new, out with the old? auto-extraction for remote sensing archaeology, in: Remote sensing of the ocean, sea ice, coastal waters, and large water regions 2012, Vol. 8532, International Society for Optics and Photonics, 2012, p. 853206.
Davis, D.S., Sanger, M.C., Lipo, C.P., 2019. Automated mound detection using lidar and object-based image analysis in beaufort county, south carolina. Southeastern Archaeology 38 (1), 23–37.
A. De Boer, Using pattern recognition to search lidar data for archeological sites, in: The World Is in Your Eyes: Proceedings of the XXXIII Computer Applications and Quantitative Methods in Archaeology Conference (March 2005–Tomar, Portugal), 2005, pp. 245–254.
Douzas, G., Bacao, F., 2018. Effective data generation for imbalanced learning using conditional generative adversarial networks. Expert Systems with applications 91, 464–471.
Dupouey, J.-L., Dambrine, E., Laffite, J.-D., Moares, C., 2002. Irreversible impact of past land use on forest soils and biodiversity. Ecology 83 (11), 2978–2984.
Ford, H.A., Walters, J.R., Cooper, C.B., Debus, S.J., Doerr, V.A., 2009. Extinction debt or habitat change?–ongoing losses of woodland birds in north-eastern new south wales, australia. Biological conservation 142 (12), 3182–3190.
Foster, D., Swanson, F., Aber, J., Burke, I., Brokaw, N., Tilman, D., Knapp, A., 2003. The importance of land-use legacies to ecology and conservation. Bioscience 53 (1), 77–88.
D. François, J. Humbert, Quantification spatiale des précipitations. applications au nord-est de la france, Revue Géographique de l'Est 40 (1–2).
Fraterrigo, J.M., Turner, M.G., Pearson, S.M., Dixon, P., 2005. Effects of past land use on spatial heterogeneity of soil nutrients in southern appalachian forests. Ecol. Monogr. 75 (2), 215–230.
Freeland, T., Heung, B., Burley, D.V., Clark, G., Knudby, A., 2016. Automated feature extraction for prospection and analysis of monumental earthworks from aerial lidar in the kingdom of tonga. J. Archaeol. Sci. 69, 64–74.
Froehlicher, L., Schwartz, D., Ertlen, D., Trautmann, M., 2016. Hedges, colluvium and lynchets along a reference toposequence (habsheim, alsace, france): history of erosion in a loess area, Quaternaire. Revue de l'Association française pour l'étude du Quaternaire 27 (2), 173–185.
Fuller, J.L., Foster, D.R., McLachlan, J.S., Drake, N., 1998. Impact of human activity on regional forest composition and dynamics in central new england. Ecosystems 1 (1), 76–95.
Gallwey, J., Eyre, M., Tonkins, M., Coggan, J., 2019. Bringing lunar lidar back down to earth: Mapping our industrial heritage through deep transfer learning. Remote Sensing 11 (17), 1994.
Georges-Leroy, M., 2020. Typologie d'après lidar des structures agraires et parcellaires fossilisées sous couvert forestier en lorraine. Archéologies numériques 1 (4), 16.
M. Georges-Leroy, A. Montes, L'apport du lidar à la reconnaissance et à l'étude des vestiges archéologiques dans le massif forestier de haye, 2013.
Grossmann, E.B., Mladenoff, D.J., 2008. Farms, fires, and forestry: disturbance legacies in the soils of the northwest wisconsin (usa) sand plain. For. Ecol. Manage. 256 (4), 827–836.
Guyot, A., Hubert-Moy, L., Lorho, T., 2018. Detecting neolithic burial mounds from lidar-derived elevation data using a multi-scale approach and machine learning techniques. Remote Sensing 10 (2), 225.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
Herrault, P.-A., Larrieu, L., Cordier, S., Gimmi, U., Lachat, T., Ouin, A., Sarthou, J.-P., Sheeren, D., 2016. Combined effects of area, connectivity, history and structural heterogeneity of woodlands on the species richness of hoverflies (diptera: Syrphidae). Landscape ecology 31 (4), 877–893.
R. Hesse, Detecting former field systems with airborne lidar–an overview of current methods.
Howard, J., et al., 2018. The fast. ai deep learning library, lessons, and tutorials.
IGN, Rge alti version 2.0 - descriptif de livraison.
Jamin, A., Peintinger, M., Gimmi, U., Holderegger, R., Bergamini, A., 2020. Evidence for a possible extinction debt in swiss wetland specialist plants. Ecology and Evolution 10 (3), 1264–1277.
Kanianska, R., 2016. Agriculture and its impact on land-use, environment, and ecosystem services. Landscape ecology-The influences of land use and anthropogenic impacts of landscape creation 1–26.
Kolk, J., Naaf, T., Wulf, M., 2017. Paying the colonization credit: converging plant species richness in ancient and post-agricultural forests in ne germany over five decades. Biodivers. Conserv. 26 (3), 735–755.
I. Kramer, An archaeological reaction to the remote sensing data explosion, Unpublished thesis (Master), University of Southampton.
J. Krauss, R. Bommarco, M. Guardiola, R. Heikkinen, A. Helm, M. Kuussaari, R. Lindborg, Habitat fragmentation, biodiversity loss and extinction debt, Ecology Letters.
Lambers, K., Verschoof-van der Vaart, W.B., Bourgeois, Q.P., 2019. Integrating remote sensing, machine learning, and citizen science in dutch archaeological prospection. Remote Sensing 11 (7), 794.
Li, W., Chen, C., Su, H., Du, Q., 2015. Local binary patterns and extreme learning machine for hyperspectral imagery classification. IEEE Trans. Geosci. Remote Sens. 53 (7), 3681–3693.
Lin, Q., Qi, W., 2015. Multi-scale local binary patterns based on path integral for texture classification. In: 2015 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 26–30.
Menze, B.H., Ur, J.A., 2012. Mapping patterns of long-term settlement in northern mesopotamia at a large scale. Proc. Nat. Acad. Sci. 109 (14), E778–E787.
Menze, B.H., Ur, J.A., Sherratt, A.G., 2006. Detection of ancient settlement mounds. Photogrammetric Engineering & Remote Sensing 72 (3), 321–327.
Niklasson, M., Granström, A., 2000. Numbers and sizes of fires: long-term spatially explicit fire history in a swedish boreal landscape. Ecology 81 (6), 1484–1499.
Ojala, T., Pietikäinen, M., Mäenpää, T., 2000. Gray scale and rotation invariant texture classification with local binary patterns. In: European Conference on Computer Vision. Springer, pp. 404–420.
Parcak, S.H., 2009. Satellite remote sensing for archaeology. Routledge.
Piessens, K., Hermy, M., 2006. Does the heathland flora in north-western belgium show an extinction debt? Biol. Conserv. 132 (3), 382–394.
Powers, J.S., 2004. Changes in soil carbon and nitrogen after contrasting land-use transitions in northeastern costa rica. Ecosystems 7 (2), 134–146.
L. Prechelt, Early stopping-but when?, in: Neural Networks: Tricks of the trade, Springer, 1998, pp. 55–69.
Pregesbauer, M., 2013. Object versus Pixel-Classification Techniques for high resolution airborne remote sensing data.
Rhemtulla, J.M., Mladenoff, D.J., Clayton, M.K., 2009. Legacies of historical land use on regional forest composition and structure in wisconsin, usa (mid-1800s–1930s–2000s). Ecological applications 19 (4), 1061–1078.
M.A. Riley, Automated detection of prehistoric conical burial mounds from lidar bare-earth digital elevation models, A thesis presented to the Department of Geology and Geography in candidacy for the degree of Master of Science, North Missouri State University Maryville, Missouri, MO, USA.
Sammut, C., Webb, G.I., 2011. Encyclopedia of machine learning. Springer Science & Business Media.
Schneider, A., Takla, M., Nicolay, A., Raab, A., Raab, T., 2015. A template-matching approach combining morphometric variables for automated mapping of charcoal kiln sites. Archaeological Prospection 22 (1), 45–62.
D. Schwartz, V. Robin, P. Adam, P. Schaeffer, A. Gebhart, P.A. Herrault, B. Keller, D. Dapiaggi, C. Stevenel, M. Thiss, et al., Les géosciences au service de l'archéologie agraire. une étude de cas sur les rideaux de culture de goldbach (68), Archimède: archéologie et histoire ancienne (7) (2020) 205–216.
Sevara, C., Pregesbauer, M., Doneus, M., Verhoeven, G., Trinks, I., 2016. Pixel versus object–a comparison of strategies for the semi-automated mapping of archaeological features using airborne laser scanning data. Journal of Archaeological Science: Reports 5, 485–498.
Smithwick, E.A., Harmon, M.E., Domingo, J.B., 2007. Changing temporal patterns of forest carbon stores and net ecosystem carbon balance: the stand to landscape transformation. Landscape Ecol. 22 (1), 77–94.
Szabo, J.K., Vesk, P.A., Baxter, P.W., Possingham, H.P., 2011. Paying the extinction debt: woodland birds in the mount lofty ranges, south australia. Emu-Austral Ornithology 111 (1), 59–70.
Toumazet, J.-P., Vautier, F., Roussel, E., Dousteyssier, B., 2017. Automatic detection of complex archaeological grazing structures using airborne laser scanning data. Journal of Archaeological Science: Reports 12, 569–579.
Trier, Ø.D., Pilø, L.H., 2012. Automatic detection of pit structures in airborne laser scanning data. Archaeological Prospection 19 (2), 103–121.
Trier, Ø.D., Zortea, M., Tonning, C., 2015. Automatic detection of mound structures in airborne laser scanning data. Journal of Archaeological Science: Reports 2, 69–79.
Trier, Ø.D., Cowley, D.C., Waldeland, A.U., 2019. Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on arran, scotland. Archaeological Prospection 26 (2), 165–175.
Vellend, M., Verheyen, K., Jacquemyn, H., Kolb, A., Van Calster, H., Peterken, G., Hermy, M., 2006. Extinction debt of forest plants persists for more than a century following habitat fragmentation. Ecology 87 (3), 542–548.
Verschoof-van, d.V.W., Lambers, K., et al., 2019. Learning to look at lidar: The use of r-cnn in the automated detection of archaeological objects in lidar data from the netherlands. Journal of Computer Applications 2, 10.

Wang, M., Deng, W., 2018. Deep visual domain adaptation: A survey. Neurocomputing 312, 135–153.
Zingman, I., Saupe, D., Penatti, O.A., Lambers, K., 2016. Detection of fragmented rectangular enclosures in very high resolution remote sensing images. IEEE Trans. Geosci. Remote Sens. 54 (8), 4580–4593.
Zuber, M.T., Smith, D.E., Zellar, R.S., Neumann, G.A., Sun, X., Katz, R.B., Kleyner, I., Matuszeski, A., McGarry, J.F., Ott, M.N., et al., 2010. The lunar reconnaissance orbiter laser ranging investigation. Space Sci. Rev. 150 (1), 63–80.