sustainability

Article

Towards Automatic Burrow Detection for Sustainable River Levees

Lisa Borgatti [1], Alberto Cervellati [2], Monica Ghirotti [3], Davide Martinucci [4], Giacomo Pampalone [5], Alberto Paparella [5], Stefano Parodi [6], Federica Pellegrini [7], Edoardo Ponsanesi [5], Guido Sciavicco [5,*], Massimo Valente [6] and Roberta Zambrini [4]

1. Department of Civil, Chemical, Environmental and Materials Engineering, University of Bologna, 40136 Bologna, Italy; lisa.borgatti@unibo.it
2. Ufficio Sicurezza Territoriale e Protezione Civile Ravenna, 48121 Ravenna, Italy
3. Department of Physics and Earth Sciences, University of Ferrara, 44121 Ferrara, Italy; monica.ghirotti@unife.it
4. Esplora S.R.L., 34123 Trieste, Italy
5. Department of Mathematics and Computer Science, University of Ferrara, 44121 Ferrara, Italy; giacomo.pampalone@edu.unife.it (G.P.); alberto.paparella@unife.it (A.P.)
6. Agenzia Interregionale per il fiume Po, 41122 Modena, Italy
7. Ufficio Sicurezza Territoriale e Protezione Civile Reggio Emilia, 42121 Reggio Emilia, Italy; federica.pellegrini@regione.emilia-romagna.it

* Correspondence: guido.sciavicco@unife.it

Abstract

Burrows are tunnels or holes excavated into the ground by certain types of animals, to be used as habitation or temporary refuge, or as a by-product of their locomotion. Burrows provide a form of shelter against predation and exposure to the elements, and can be found in nearly every biome and among various biological interaction types. River bank burrowing weakens the soil structure, increases the risk of erosion, and may lead to bank retreat and landslides. Currently, burrow watching, mapping, and prevention are human-only activities, and there are no conventional data or information systems designed for this purpose. In this paper, we design, implement, and test a novel AI-based solution that, starting with drone-acquired imagery, allows the user to automatically identify and map potentially dangerous burrows in the target area, and lays the basis for the digitization and systematic conservation of such information, to be later used for intervention and planning. Our solution contributes to the environmental sustainability of rivers, especially close to densely populated areas.

Keywords: artificial intelligence; remote sensing; burrows; levees; flood protection; sustainable environment; ecosystem; environmental governance; environmental management

Academic Editor: Faccini Francesco

Received: 15 December 2025
Revised: 23 January 2026
Accepted: 6 February 2026
Published: 23 February 2026

1. Introduction

River embankments serve as critical infrastructure for flood protection, safeguarding communities and agricultural lands from catastrophic inundation. However, the progressive deterioration, damage and eventual breach of river embankments can result from several failure scenarios or chain of events. These failures are typically driven by processes such as internal erosion and instability affecting both the waterside and landside slopes. The performance of river embankments is significantly influenced by the presence of such macro pores in the ground, which contribute to the progressive degradation of soil properties. These discontinuities negatively affect water pressure distribution, diminishing both effective stresses and shear strength, ultimately promoting instability. Furthermore, cavities breaching the levee's full width can rapidly accelerate internal erosion, leading to catastrophic failure [1,2].

Among the various macro pores, animal burrows stand out as a primary source of biological erosion in earthen structures like levees and dams. The consequences of wildlife activities in levees can be categorized into structural damage, surface erosion and hydraulic alterations; impacts that are often interrelated and mutually reinforcing. Hydraulic performance changes are one of the most common causes of failure in earthen structures, which typically lead to internal erosion and piping [3], a phenomenon responsible for nearly half of global dam failures [4]. Research has demonstrated that burrowing activity severely compromises levee stability through mechanisms that can induce complete failure and catastrophic flooding [5–7]. Studies show how the hydraulic impact of rising water levels on earthen levees, occurring during heavy and prolonged periods of rain, is severely compounded in damaged levees, leading to an increased risk of failure. In such a state, the conditions for the onset of inner erosion are achieved for shorter flood durations [8]. Given the increasing risks associated with animal burrowing activity, a threat potentially amplified by changing precipitation patterns, altered wildlife behavior and ecosystem shifts, the development of efficient methodologies for early cavity identification is essential for effective risk mitigation and adaptive infrastructure management. Prior efforts in this domain have largely focused on geophysical techniques. These methods have been widely applied to levees, primarily to image the internal structure and estimate hydrogeological features [9]. Other studies have specifically used techniques like resistivity, magnetic measures [10] and Ground Penetrating Radar (GPR) [11–13] for the direct detection of internal voids and burrows. However, these approaches typically require specialized equipment, trained personnel and significant time investment, limiting their application for continuous, system-wide monitoring.

Our work proposes a data-driven approach that enables scalable, resource-efficient surveillance of extensive levee networks. By applying machine learning techniques to remotely sensed imagery, we can automatically identify and map surface patterns consistent with burrows, enabling continuous surveillance at a fraction of the cost, resource requirements and environmental footprint of traditional survey methods. A machine learning model was dynamically trained on real-world examples via a dedicated platform, progressively improving detection accuracy and adapting to regional variations in wildlife activity and embankment characteristics. This system offers precise identification of threats and provides an efficient and accurate way to manage and monitor the territory. This approach supports a fundamental shift from reactive, resource-intensive emergency repairs to proactive maintenance strategies that extend infrastructure lifespan, optimize resource allocation and minimize the environmental disruption associated with catastrophic failures. Beyond detection, our objective is to furnish an innovative tool for adaptive land management that provides precise threat identification, efficient territorial monitoring and comprehensive data collection capabilities, enabling more effective and timely interventions that balance infrastructure protection with broader ecosystem considerations. In a single sentence, we want to answer the question:

Is it possible to automatically identify animal burrows in riverbanks?

Our system is fully implemented and already at the disposal of regional officials, who contribute to its improvement and test by using it as part of their normal activities.

Automatic detection of wildlife activities in levees is extremely relevant for the sustainability of the environment at the local level. Progressively shifting the monitoring load and its responsibility from human-based to machine-based systems could enable more restrictive and safer policies without increasing the economic cost for local and re-

https://doi.org/10.3390/su18042153

gional communities, therefore allowing a prompter response to the ever more pressing environmental changes.

2. Related Works

Manual inspection of river levees and embankments is not only time consuming, often taking hours to days depending on the complexity and amount of cases, but also prone to human error, particularly when identifying subtle anomalies or working under time constraints. This has motivated significant research into both geophysical detection methods and automated visual inspection systems for identifying structural vulnerabilities in earthen structures.

Several non-destructive methods are available to detect and map underground voids [9,10,14], with applications ranging from cavity detection in earthen structures to archeological surveys. Among these methods are gravity survey, resistivity methods, seismic reflection and Ground Penetrating Radar (GPR) [11–13]. Gravity survey is a non-destructive geophysical technique that measures differences in the Earth's gravitational field at specific locations. Its success depends on identifying variations in the measured gravitational field due to differences in the bulk density of earth materials. This method is considered one of the most cost-effective geophysical approaches when investigating wide areas with relatively large caverns [15–17]. Resistivity methods have also been applied for burrow detection, exploiting the fact that the electrical resistance of air filled voids is higher than the surrounding soils [18]. Since air filled voids provide an excellent dielectric constant contrast, GPR has been successfully used to identify animal burrows in earthen structures [19]. GPR surveys have yielded reliable information about lithology and internal structure of river banks, including the identification of flood terraces and detection of damage failures due to water percolation, out-washing and erosion. Kinlaw et al. [20] and Cortez et al. [21] used GPR to visually map underground burrows of the Gopher Tortoise. Field studies utilizing GPR to monitor river embankments have been reported by Di Prinzio et al. [22]. For more in-depth discussions on the presence of animal burrows, their impact and management, we refer to Bayoumi et al. [3] and the Federal Emergency Management Agency report [1].

Despite their accuracy, geophysical techniques, and, in particular, GPR, still have penetration limitations, especially in conductive silty-clayey soils (1.5 m). They also require specialized equipment, qualified personnel and a significant investment of time, which limits their feasibility for continuous, large scale surveys and monitoring. These limitations have motivated research into complementary automated visual detection methods that can rapidly process imagery from routine inspections or aerial surveys.

The automated detection of burrows in levee imagery exemplifies a broader class of computer vision problems that require the identification of small scale features within large natural images. The general problem can be formulated as follows: given a collection of high resolution images containing sparse instances of target features embedded within complex, heterogeneous backgrounds, develop an automated system capable of locating and classifying these features with sufficient reliability to support operational decision making. Similar challenges have been addressed in other environmental monitoring domains. Lieshout et al. [23] and Hidaka et al. [24] employed Convolutional Neural Networks (CNNs) to detect and quantify floating plastic pollution in rivers using deep learning and camera imagery. The latter specifically proposes and applies semantic segmentation, a pixel-based classification approach to recognize and distinguish between natural and artificial litter along sea shores. This methodology has been applied to images captured by human operators via Unmanned Aerial Systems (i.e., drones), an approach further elaborated by Gonçalves et al. [25] and Kako et al. [26] as a low cost monitoring technique.

Various authors propose aerial surveys as a solution to monitor coasts and shores and identify plastic waste [27–30]. These studies share key similarities with burrow detection: both involve identifying relatively small, visually subtle features within complex natural environments; both benefit from aerial or elevated imaging perspectives and require distinguishing target features from highly variable background textures including vegetation, soil, water and shadow patterns.

In the field of anomaly detection, Nayak et al. [31] and Sifnaios et al. [32] suggest a pixel-first approach, building up from independent pixel-level classifications and leading to frame-level classification, using data such as the quantity of detected anomalous pixels and their spatial distribution. This approach has proven particularly effective in medical imaging, where image segmentation techniques have been crucial in diagnosing diseases for many years, transforming images into clear and visually interpretable sections that facilitate recognition [33]. Random Forest classifiers [34] have proven particularly effective for environmental monitoring tasks involving heterogeneous natural imagery. This method combines multiple decision trees to improve classification robustness and reduce overfitting, all critical advantages when dealing with the inherent variability in field imagery. Unlike deep learning approaches that require large labeled datasets and significant computational resources, Random Forests offer several practical advantages: they handle high-dimensional feature spaces efficiently, provide interpretable feature importance rankings, generate probabilistic confidence estimates, and can achieve strong performance with relatively modest training datasets. Furthermore, various studies support their dominance over deep learning when working with tabular data [35–38]. Random Forests have been successfully applied to land cover classification [39–41], vegetation mapping [42] and detection of environmental anomalies in remote sensing imagery [43]. Their ability to assess feature importance is particularly valuable for understanding which visual characteristics (i.e., texture, color, shape, and context) are most discriminative for target detection.

3. Materials and Methods

After briefly introducing the study site, in this section, we describe the hardware and software used in the image collection process, detailing its context and the distinctly applied procedures, as well as the characteristics of the obtained results. Then, we describe the general workflow covering the differences between traditional monitoring missions and our contribution's approach. Lastly, we present the pipeline of our methodology divided into three phases: gathering of user-provided labels, the training of a first filtering model and the final detection of burrow-like structures whose location will be reported on an associated map relative to this specific monitoring campaign.

3.1. Study Area

Our proposed methodology has been tested on images extracted from a surveillance campaign conducted along selected reaches of the Senio River, a 92 km tributary of the Reno River in the Emilia-Romagna region of Italy (44.5475° N, 12.0954° E). The Senio originates in the Appennino Tosco-Emiliano mountains in the province of Florence and flows northeast through the province of Ravenna, before its confluence with the Reno. The river exhibits highly variable discharge patterns, with a mean flow of approximately 10 m^3/s, ranging from minimum flows of 0.3 m^3/s during dry periods to peak discharges exceeding 500 m^3/s during flood events. Notably, in May 2023 the river experienced significant flooding as part of the widespread inundation events that affected the Emilia-Romagna region. The selected areas represent sections where animal burrowing activity posed potential concerns for bank integrity and flood risk management.

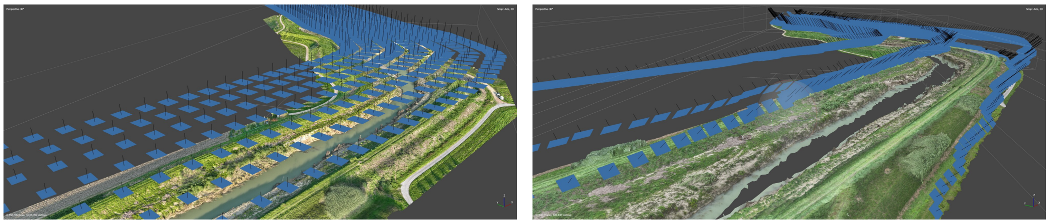
3.2. Image Collection

Aerial imagery was acquired using an Unmanned Aircraft System (UAS) platform equipped with a DJI Mavic 3 Enterprise produced by DJI Technology (Shenzhen, China) and a 20-megapixel RGB camera (Shenzhen, China). The drone was integrated with an RTK module connected to the Hexagon SmartNet network, which provided real-time GNSS correction services to achieve centimeter accuracy (horizontal: 1 cm; vertical: 1.5 cm + 1 ppm) for camera positions and general navigation. Choosing RBG-only imagery was motivated by economic costs, as well as the driving principle of building a system that works based on the same premises as human-assisted identification. The camera featured a three axis stabilized RGB sensor with a wide-angle lens (24 mm equivalent focal length, f/2.8–f/11 aperture) and a 4/3″ CMOS sensor with 3.3 µm pixels (maximum image dimensions: 5280 × 3956 px). The system was chosen for its stability in variable wind conditions, typical of riparian corridors and its capability to capture high resolution imagery, suitable for detecting small scale geomorphological features such as burrow entrances. Prior to field deployment, flight missions were planned using DJI Pilot 2 to ensure systematic coverage of target riverbank sections. Flight plans were designed to capture both banks of the river through separate parallel flight lines to better capture the details along the levee. Survey altitude was maintained at approximately 30–35 m above ground level (AGL), providing a ground sampling distance of approximately 9 mm/pixel. This resolution was selected to enable reliable identification and measurement of burrow openings while maintaining efficient coverage of extended river reaches. The collected data were georeferenced to the RDN2008/UTM 32N coordinate system (North-East coding, EPSG Code 6707). Elevation values are based on the ellipsoidal height system. Image overlap was configured at approximately 80% longitudinal overlap and 60% lateral overlap to ensure adequate redundancy for subsequent photogrammetric processing and to minimize the risk of data gaps. The UAS was flown at a consistent speed below 3 m/s to maintain image quality and reduce motion blur. Nadir (vertical) imagery was acquired in automated flight mode for general coverage and photogrammetric model construction, while more detailed oblique shots of burrows along the internal and external sides of the river banks were captured through manual flight with time interval programmed capture. The positional accuracy (RMSE) of camera positions ranged between 1.5 cm and 2.5 cm. In total, approximately 0.155 km^2 were surveyed along a 950 m linear river reach. Flight operations were conducted during the autumn season (2024) to take advantage of optimal conditions, such as low water levels for maximum bank exposure and favorable vegetation conditions. Missions were scheduled during midday hours to minimize shadow effects and ensure consistent illumination across the survey area. All flights were performed under suitable meteorological conditions. No special authorizations were required for the flight operations, though nearby landowners and residents were notified prior to survey activities. Concurrent with UAS operations, ground-based topographic surveys of burrow entrances and feeding holes were conducted using a Stonex S990a GNSS topographic system (Stonex, Paderno Dugnano, Italy) operating in nRTK mode and connected to the Hexagon SmartNet network for real-time data correction. Having this fundamental ground truth allowed objective verification of the number and precise locations of these features. The image acquisition methodology was subject to several constraints inherent to UAS-based surveys of riparian environments. Examples of the shots can be seen in Figure 1, while in Figure 2 provides an overview of all of the taken shots. Dense vegetation in certain reaches limited visibility of bank faces, especially in areas lacking recent maintenance. Additionally, the single date nature of the survey represents a snapshot of conditions and does not capture the temporal dynamics of burrow activity, nor seasonal variations in detectability.

https://doi.org/10.3390/su18042153

(**a**) (**b**) (**c**) (**d**)

Figure 1. Examples of collected images: (**a**,**b**) show a nadiral view of the river, figures (**c**,**d**) show the specific details of burrows along the levee.



(**a**) (**b**)

Figure 2. Overview of all the shots taken via drone on the site: figure (**a**) refers to the nadiral shots taken during the programmed flight, while figure (**b**) refers to the inclined shots of the levee taken with human supervision.

3.3. Workflow

After acquisition, images are stored and organized into a hierarchical structure, encompassing the levels of campaigns and image sets. A campaign includes all images from a specific monitoring initiative conducted at a particular site. Each campaign is further subdivided into multiple image sets, grouping images based on shared characteristics such as shooting position, angle and environmental conditions. The entire campaign is stored in a centralized database. Traditionally, this workflow would require users to manually label images and identify the areas of the riverbank that are deemed most at risk, a time consuming process that demands expert knowledge and careful inspection. The classification workflow aims at automating this critical step through a systematic approach: users manually label representative image regions to identify burrow and non-burrow areas, establishing training examples. To enable effective automation, it is essential to extract general features that can reliably distinguish burrows across diverse conditions. These annotations enable the training of a pixel-level classification model that learns to distinguish patterns associated with burrows from background elements. Once trained, the model processes complete images to generate binary classification outputs, which then undergo morphological filtering and geometric analysis to identify any candidate burrow locations. Results are presented to users for validation and correction, thus enabling an iterative refinement of the model's performance. The classification pipeline is illustrated in Figure 3.

The implementation of this workflow faces significant challenges inherent to the nature of riverbank monitoring data. Image reconnaissance is demanding in resources and time consuming, often requiring specialized equipment and favorable environmental conditions. This constraint frequently results in insufficient training examples to adequately cover the full variance of burrow characteristics, including shape, size and color across various contexts. Variations across geographical zones, like vegetation density, burrow dimensions, illumination conditions (e.g., due to the time of the day), image inclination and other environmental factors substantially increase the complexity of the prediction problem. Traditional monitoring campaigns are inherently constrained by temporal and environmental factors, hence typically capturing images during optimal lighting conditions within specific seasonal windows. This results in the available training data rarely covering the full spectrum of possible burrow appearances across different geographical, climatic, seasonal and meteorological conditions. These sources of variability affect not only the visual appearance of burrows, but also the background context, making feature extraction and model generalization particularly challenging. Consequently, developing a single general-purpose model capable of reliably detecting burrows across all possible conditions proves impractical.

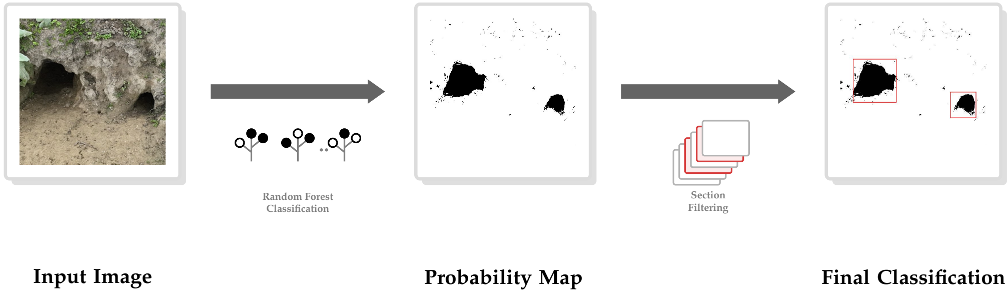


Figure 3. The classification pipeline: The first layer is composed of a random forest that provides a binary classification for each pixel, while the second layer makes use of density and shape properties inside of the binary frame, allowing for accurate classification of the whole frame.

To address these challenges, rather than pursuing a single universal predictive model, we implemented a flexible, context-dependent modeling strategy. This approach enables the development of customized models specifically tailored to the characteristics of each image set within its respective campaign. By training lightweight models adapted to the specific visual conditions of each image set, the system can effectively handle the high variability observed in field data while maintaining computational efficiency. This strategy balances automation with expert oversight, allowing the system to progressively refine its detection capabilities while maintaining flexibility across diverse environmental conditions and image characteristics. The modular design also supports model reuse across visually similar image sets, reducing redundant annotation efforts and enabling efficient scaling across multiple monitoring sites. The process can be divided into three phases: manual annotation, model training, and frame classification and refinement. These will be detailed in the following sections.

3.4. Manual Annotation

In the manual annotation phase, the user manually labels regions within each image, distinguishing between positive samples (i.e., corresponding to burrows) and negative samples (i.e., non-burrow areas). This task is carried out through a dedicated graphical interface, which allows to inspect images visually, zoom-in and pan as needed. Using the mouse, the user draws rectangular regions of variable size and explicitly assigns each one to a category, either burrow or non-burrow. The system automatically records the selection's coordinates, width and height. Each annotated region is stored while maintaining its association with the corresponding image for subsequent analysis and model development. In this process, identifying zones free of burrow activity is critically important, as these negative samples play a key role in defining a clear distinction between the two categories. By providing examples of areas where burrows are absent, the system can better distinguish relevant features and reduce the likelihood of false positives during training, thus proficiently ruling out noise and inference. This manual annotation defines the foundational knowledge that the predictive system will learn from. The overall analysis is based on a continuous feedback loop between the user and the application, in which the human component plays a fundamental role in conveying to the system the concept of a burrow, guiding it in the discrimination between positive and negative areas.

3.5. Model Training

In the model training phase, a machine learning algorithm uses the labeled regions from the first phase to generate a predictive model capable of distinguishing between burrow and non-burrow pixels. The model relies primarily on the color information of individual pixels (RGB values) without explicitly considering spatial relationships between neighboring pixels. Specifically, the algorithm trains a decision tree ensemble, namely a Random Forest, which combines the output of multiple decision trees to improve classification accuracy and robustness. The transformation from image regions to individual pixel classifications allows the application of a Random Forest to this computer vision task. Each pixel within the selected positive and negative image regions was treated as an independent classification instance, with its RGB color values serving as the three-dimensional feature vector input to the machine learning algorithm.

The pixel classification process systematically sampled every pixel within the designated positive and negative regions, creating a comprehensive representation of the color characteristics associated with each class. For burrow regions, this approach mostly captured the dark internal pixels typically associated with cavity openings, but also the transitional pixels at the edges of the burrow where shadow gradients create intermediate color values. Similarly, negative regions comprised pixels that cover the full range of background color variations, from bright soil surfaces to shadowed areas that might be erroneously misclassified as actual burrow openings. Once trained, the Random Forest model was applied to classify every pixel within new target images requiring burrow detection.

Each pixel in these candidate images was extracted and evaluated by the classifier, which assigned it to either the burrow or non-burrow class based on its color characteristics. The extraction and training process is represented in Figure 4. This pixel-by-pixel classification produces a binary output frame, a black and white representation where black pixels indicate locations classified as potential burrow regions, and white pixels encapsulate multiple non-burrow elements, such as background areas (sky, water, etc.), vegetative elements (plants and trees) or artificial structures (roads and signs).

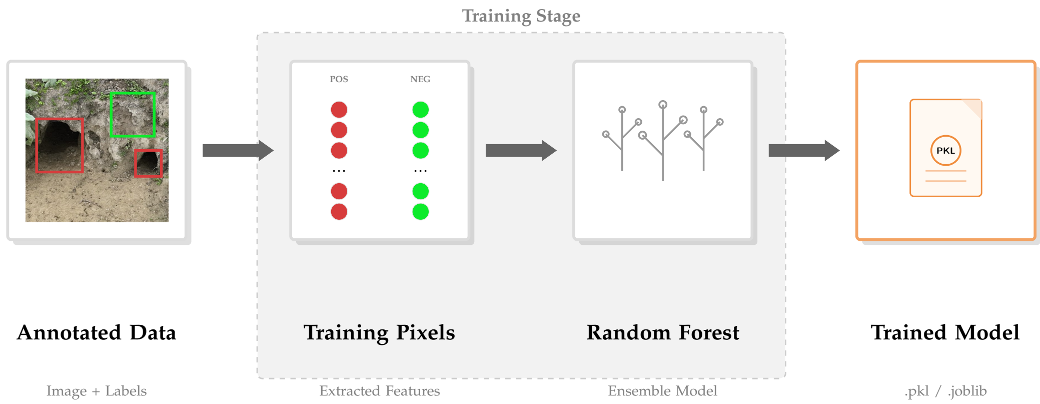


Figure 4. The training process pipeline. Pixels are extracted from user-selected examples to train a Random Forest model able to recognize the differences between positive and negative pixels.

This binary frame represents an intermediate processing stage rather than a final detection result. While it successfully highlights candidate regions where burrow-like color patterns exist, the pixel-level classification alone does not account for spatial coherence, morphological characteristics, nor any contextual features that distinguish true burrow openings from false positives. The binary frame may contain scattered individual pixels or small clusters classified as burrows due to color similarity. Consequently, this initial classification output requires further post-processing steps, including morphological operations for noise reduction, contour detection and feature-based filtering to consolidate fragmented detections, eliminate spurious classifications and extract meaningful burrow candidates for subsequent validation and measurement.

The performance of the predictive model strongly depends on the quality and precision of the manual annotations from the first phase. Accurate labeling of burrow and non-burrow regions ensures that the system can learn reliable visual patterns and environmental cues, ultimately improving detection accuracy across diverse images and scenarios.

3.6. Frame Classification

Finally, during classification and refinement, the trained model is deployed across the entire image dataset to automatically detect and localize burrows. Following a cleaning step, the method additionally filters sub-areas out of the black and white binary frame, selecting those that contain certain morphological and spatial properties typical of a visible burrow opening. Three low-level image cleaning steps are performed at this point: noise reduction, consisting of cleaning and removing the sporadic false positive pixels; the application of a density filter, that is, letting a sliding window through the binary frame, by means of which, given the density ratio of the positive burrow and total pixels in each window, we extract the windows where that ratio surpasses a certain threshold; lastly, the application of a geometric filter allows us to search and extract agglomerates of positively classified pixels that can be inscribed inside of a circle, evaluating the objects' circularity.

The model outputs are then presented back to the user, who can assess their accuracy and, if necessary, refine or extend the existing labels by correcting erroneous classifications. These corrections are incorporated into the training dataset and the cycle repeats. This iterative process of learning and correction allows the system to progressively improve its performance and ability to generalize, with each iteration refining the model's understanding of what constitutes a burrow. Each trained model is therefore explicitly linked to the image set from which it was generated. This design choice provides several advantages. Firstly, it ensures that users can rely on stable and well-performing models for specific portions of their data (i.e., images), independent of the variability present in the rest of the dataset. Secondly, it allows for the reuse of similar models. Through comparative analysis, users may identify multiple image sets that share visual or contextual similarities. In such cases, instead of training a new model from scratch (including the entire manual labeling process), users can test previously trained models on new image sets to evaluate their flexibility, applicability and general performance. This mechanism promotes efficiency, reduces computational costs and encourages the incremental refinement of predictive models based on empirical feedback.

4. Experiments and Results

To validate the efficacy of the proposed methodology for animal burrow detection in river levee aerial imagery, we conducted a comprehensive experimental evaluation. The assessment framework was designed to quantify performance at two distinct levels: the pixel-wise classification capability of the initial Random Forest stage and, crucially, the spatial accuracy of the final detection system in localizing actual burrow instances.

The experimental dataset comprises selected aerial images from the Senio River dataset, that capture levee infrastructure and animal burrows with high spatial detail. These images serve as the testing ground for the pipeline, encompassing a representative variety of environmental conditions.

4.1. Dataset

The images collected are described in Section 3.2 and are characterized by high resolution (5280 × 3956 px) and substantial overlap between each subsequent shot, resulting in

https://doi.org/10.3390/su18042153

broad areas encompassing numerous burrows. Because of the specific characteristics of the area and of the animals that populate it, the set of images does not contain negative frames; that is, burrows occur in every image (see Section 4.4). Thus, we processed the images as follows: the image set was linearly spaced and each image was divided into six equally sized quadrants, 52 containing burrows and 35 without any burrows, resulting in a total of 87 images. Figure 5 shows examples of the extracted dataset snippets, with their relative manual classification.

(a) (b) (c) (d)

Figure 5. Examples of images extracted for the dataset: (**a**,**b**) show images without burrows, while (**c**,**d**) show images with burrows.

4.2. Labeling

The objective assessment of the automated classification results relies on a comparison against the labeled dataset, which serves as the authoritative reference standard. In this study, ground truth was established through a rigorous manual annotation process. Human experts meticulously inspected the dataset to classify regions of interest. Areas confirmed to exhibit burrow characteristics were manually delineated and labeled as *positive* instances, creating a definitive classification mask. The resulting set of labeled instances provides the baseline required to calculate the performance metrics detailed in the following sections.

Consistency in the manual annotation phase was guaranteed by having several independent experts contributing to this step; the area under analysis was already known to each expert, who had several years of experience in this field.

4.3. Experimental Protocol

To rigorously assess the robustness and the zero-configuration capability of the proposed method, we designed a comprehensive experimental protocol. The global optimization was conducted by a full grid search strategy, exploring the interplay between the morphological filtering and spatial density parameters. We evaluated every possible combination within the defined parameter space to capture interactions between the morphological cleaning intensity and the detection criteria, in particular, we combined the following: five morphological configurations ($C \in \{0, 1, 2, 3, 4\}$) representing a range of morphological filtering intensity, from raw segmentation to aggressive noise reduction, and eight density thresholds ($D \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$), encompassing a range of thresholds defining the required proportion of positive pixels within a candidate section.

The parameters C and D are intrinsically coupled. Morphological cleaning acts by eroding the predicted positive areas in the probability map, so the density threshold could be significantly affected by the resulting change in the spatial distribution and quantity of these pixels. By mapping the entire performance landscape, we can assess the stability of the proposed descriptors. A robust method should exhibit a wide plateau of high performance rather than a single, isolated peak. To measure the impact of data availability and perceived effort for the user, the entire grid search was repeated varying the training set size ($size \in \{5\%, 10\%, 20\%, 40\%, 60\%, 80\%, 100\%\}$ of the training pool). This multidimensional approach allows us to quantify the learning convergence of the morphological features and demonstrates the system's ability to achieve high performance, even with limited labeled data. To ensure consistent benchmarking, we adopted a fixed hold-out validation strategy with bootstrapped training sets. Thus, prior to the experiments a dedicated test set (comprising 20% of the dataset) was isolated and kept fixed throughout all iterations; the remaining images constituted the training pool. For every experimental condition (i.e., specific configuration, density, and training size), we cross-validated the results with a 10-fold repetition of randomly sampling a subset of distinct images from the training pool according to the chosen percentage, training the model and evaluating the model on the test set. For every recorded performance index, the final reported performance is given in terms of mean value and standard deviation over the 10 runs. This approach isolates the variability due to training data selection, providing a robust estimate of the method's stability.

4.4. Evaluation Metrics

The performance evaluation is conducted at the frame (or image) level. In this context, the problem is formalized as a binary classification task: distinguishing between frames containing at least one target (positive frames) and frames containing exclusively background (negative frames). Accordingly, the metrics are defined as follows: a true positive is recorded when the system detects at least one burrow candidate in a frame that actually contains one, and a false positive is recorded when the system flags a frame as containing a burrow, but the frame is actually empty (pure background). Given the potential imbalance between frames with and without burrows, we chose to record, for every run, precision and F1-score: precision assesses the reliability of the system's alerts, and answers the question "*given that the system claims an image contains a burrow, how likely is that it is correct?*"; F1-score, defined as the harmonic mean of precision and recall, allows us to establish when high precision is not achieved simply by ignoring difficult cases, therefore summarizing the system's ability to correctly identify positive frames without generating excessive false alarms on empty ones. The results are shown in Table 1.

4.5. Discussion

Upon reading the results, the following considerations emerge. The observed invariance in both mean performance and standard deviation with minimal morphological cleaning ($C \approx 0$) and permissive density thresholds ($D \leq 0.3$) suggests a state of underfiltering saturation. In such configurations, the system essentially operates at a performance floor where the post-processing logic (cleaning and density threshold) is too loose to reflect any incremental improvements in the underlying decision tree ensemble classifier. Because the detection criteria are so easily satisfied, even a minimal 5 training set appears sufficient to generate a probability map capable of triggering a detection, effectively masking the benefits that a larger training volume would typically provide to the model's discriminative power. This lack of response to increased data is further compounded by the dominance of structural noise, such as shadows and rocky textures, which are consistently present throughout the aerial imagery. Finally, the remarkable stability of the standard deviation values across iterations is proof of statistical determinism. In more restrictive configurations, the selection of different training samples during bootstrapping typically introduces variance in the results.

https://doi.org/10.3390/su18042153

Table 1. Full Grid Search Results. Precision (P) and F1-Score (F1) metrics ($\mu \pm \sigma$). Vertical labels indicate the Morphological Filtering configuration (C), columns indicate Density Threshold (D).

		D = 0.2		**D = 0.3**		**D = 0.4**		**D = 0.5**		**D = 0.6**		**D = 0.7**		**D = 0.8**		**D = 0.9**	
	Train	**P**	**F1**	**P**	**F1**	**P**	**F1**	**P**	**F1**	**P**	**F1**	**P**	**F1**	**P**	**F1**	**P**	**F1**
C	5%	0.600 ± 0.000	0.750 ± 0.000	0.608 ± 0.014	0.756 ± 0.011	0.625 ± 0.000	0.769 ± 0.000	0.625 ± 0.000	0.769 ± 0.000	0.625 ± 0.000	0.692 ± 0.000	0.667 ± 0.072	0.799 ± 0.051	0.667 ± 0.072	0.799 ± 0.051	0.689 ± 0.054	0.793 ± 0.059
	10%	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.617 ± 0.014	0.763 ± 0.011	0.645 ± 0.041	0.783 ± 0.031	0.674 ± 0.067	0.771 ± 0.085	0.761 ± 0.068	0.750 ± 0.096
	20%	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.608 ± 0.014	0.756 ± 0.011	0.653 ± 0.028	0.790 ± 0.021	0.706 ± 0.050	0.827 ± 0.034	0.791 ± 0.112	0.863 ± 0.081	0.910 ± 0.018	0.780 ± 0.100
	40%	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.608 ± 0.014	0.756 ± 0.011	0.608 ± 0.014	0.756 ± 0.011	0.636 ± 0.042	0.777 ± 0.031	0.724 ± 0.062	0.831 ± 0.052	0.791 ± 0.117	0.810 ± 0.042	1.00 ± 0.000	0.694 ± 0.057
	60%	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.608 ± 0.014	0.756 ± 0.011	0.682 ± 0.000	0.811 ± 0.000	0.786 ± 0.074	0.861 ± 0.056	0.874 ± 0.152	0.734 ± 0.090
	80%	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.608 ± 0.014	0.756 ± 0.011	0.683 ± 0.031	0.811 ± 0.022	0.751 ± 0.038	0.858 ± 0.025	0.869 ± 0.150	0.772 ± 0.066
	100%	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.600 ± 0.000	0.750 ± 0.000	0.672 ± 0.017	0.804 ± 0.012	0.776 ± 0.023	0.874 ± 0.015	1.00 ± 0.000	0.750 ± 0.000
C = 1	5%	0.634 ± 0.016	0.776 ± 0.012	0.634 ± 0.016	0.776 ± 0.012	0.644 ± 0.033	0.783 ± 0.024	0.664 ± 0.046	0.797 ± 0.033	0.686 ± 0.063	0.797 ± 0.052	0.705 ± 0.039	0.811 ± 0.046	0.695 ± 0.051	0.788 ± 0.061	0.662 ± 0.141	0.681 ± 0.206
	10%	0.617 ± 0.014	0.763 ± 0.011	0.653 ± 0.028	0.790 ± 0.021	0.696 ± 0.064	0.820 ± 0.046	0.722 ± 0.086	0.836 ± 0.059	0.756 ± 0.038	0.843 ± 0.036	0.825 ± 0.075	0.817 ± 0.085	0.913 ± 0.088	0.754 ± 0.161	0.944 ± 0.096	0.502 ± 0.183
	20%	0.662 ± 0.017	0.797 ± 0.012	0.694 ± 0.036	0.819 ± 0.025	0.763 ± 0.023	0.866 ± 0.015	0.833 ± 0.036	0.860 ± 0.037	0.927 ± 0.072	0.858 ± 0.031	0.928 ± 0.065	0.726 ± 0.103	0.928 ± 0.065	0.671 ± 0.086	0.944 ± 0.096	0.538 ± 0.086
	40%	0.653 ± 0.028	0.790 ± 0.021	0.729 ± 0.055	0.842 ± 0.037	0.794 ± 0.069	0.884 ± 0.044	0.838 ± 0.042	0.859 ± 0.028	0.927 ± 0.072	0.844 ± 0.015	0.911 ± 0.084	0.719 ± 0.023	0.958 ± 0.072	0.606 ± 0.033	0.933 ± 0.116	0.414 ± 0.012
	60%	0.635 ± 0.030	0.776 ± 0.023	0.683 ± 0.031	0.811 ± 0.022	0.740 ± 0.055	0.850 ± 0.036	0.808 ± 0.109	0.860 ± 0.047	0.878 ± 0.113	0.881 ± 0.074	0.880 ± 0.125	0.747 ± 0.071	0.881 ± 0.206	0.631 ± 0.009	0.889 ± 0.193	0.555 ± 0.049
	80%	0.608 ± 0.014	0.756 ± 0.011	0.672 ± 0.017	0.804 ± 0.012	0.727 ± 0.039	0.842 ± 0.027	0.810 ± 0.086	0.893 ± 0.054	0.801 ± 0.065	0.830 ± 0.034	0.941 ± 0.052	0.822 ± 0.090	0.933 ± 0.116	0.638 ± 0.002	1.00 ± 0.000	0.474 ± 0.046
	100%	0.617 ± 0.014	0.763 ± 0.011	0.672 ± 0.017	0.804 ± 0.012	0.763 ± 0.023	0.866 ± 0.015	0.791 ± 0.042	0.883 ± 0.026	0.925 ± 0.003	0.870 ± 0.023	1.00 ± 0.000	0.800 ± 0.000	1.00 ± 0.000	0.636 ± 0.000	1.00 ± 0.000	0.524 ± 0.041
C = 2	5%	0.634 ± 0.016	0.776 ± 0.012	0.653 ± 0.028	0.790 ± 0.021	0.685 ± 0.052	0.812 ± 0.037	0.692 ± 0.063	0.809 ± 0.045	0.694 ± 0.048	0.787 ± 0.066	0.710 ± 0.042	0.806 ± 0.069	0.643 ± 0.096	0.658 ± 0.183	0.607 ± 0.148	0.563 ± 0.206
	10%	0.683 ± 0.031	0.811 ± 0.022	0.718 ± 0.062	0.835 ± 0.041	0.751 ± 0.038	0.858 ± 0.025	0.807 ± 0.050	0.839 ± 0.047	0.882 ± 0.055	0.825 ± 0.140	0.927 ± 0.072	0.752 ± 0.157	0.911 ± 0.078	0.551 ± 0.055	0.917 ± 0.144	0.328 ± 0.010
	20%	0.738 ± 0.021	0.849 ± 0.014	0.812 ± 0.030	0.878 ± 0.030	0.896 ± 0.120	0.853 ± 0.038	0.972 ± 0.048	0.754 ± 0.060	0.936 ± 0.055	0.763 ± 0.040	0.919 ± 0.073	0.634 ± 0.087	1.00 ± 0.000	0.514 ± 0.160	1.00 ± 0.000	0.389 ± 0.096
	40%	0.738 ± 0.021	0.849 ± 0.014	0.847 ± 0.024	0.907 ± 0.003	0.853 ± 0.061	0.857 ± 0.018	0.949 ± 0.089	0.840 ± 0.052	0.899 ± 0.010	0.719 ± 0.051	0.952 ± 0.083	0.604 ± 0.080	0.952 ± 0.083	0.513 ± 0.080	1.00 ± 0.000	0.392 ± 0.051
	60%	0.717 ± 0.057	0.835 ± 0.039	0.780 ± 0.085	0.865 ± 0.050	0.871 ± 0.105	0.896 ± 0.038	0.894 ± 0.119	0.825 ± 0.019	0.878 ± 0.135	0.716 ± 0.019	0.909 ± 0.158	0.608 ± 0.033	0.917 ± 0.144	0.603 ± 0.055	0.905 ± 0.165	0.432 ± 0.019
	80%	0.740 ± 0.055	0.850 ± 0.036	0.819 ± 0.066	0.890 ± 0.042	0.853 ± 0.094	0.898 ± 0.038	0.930 ± 0.067	0.871 ± 0.016	0.967 ± 0.058	0.740 ± 0.017	0.917 ± 0.144	0.647 ± 0.018	0.933 ± 0.116	0.594 ± 0.040	0.944 ± 0.096	0.439 ± 0.032
	100%	0.715 ± 0.034	0.834 ± 0.023	0.835 ± 0.047	0.910 ± 0.028	0.875 ± 0.000	0.903 ± 0.000	0.923 ± 0.000	0.857 ± 0.000	1.00 ± 0.000	0.783 ± 0.029	1.00 ± 0.000	0.636 ± 0.000	1.00 ± 0.000	0.571 ± 0.000	1.00 ± 0.000	0.421 ± 0.000
C = 3	5%	0.685 ± 0.052	0.812 ± 0.037	0.683 ± 0.031	0.811 ± 0.022	0.692 ± 0.063	0.809 ± 0.045	0.705 ± 0.039	0.811 ± 0.046	0.698 ± 0.054	0.759 ± 0.123	0.720 ± 0.113	0.726 ± 0.215	0.688 ± 0.055	0.654 ± 0.140	0.670 ± 0.160	0.540 ± 0.222
	10%	0.758 ± 0.095	0.860 ± 0.063	0.788 ± 0.130	0.867 ± 0.062	0.860 ± 0.065	0.867 ± 0.046	0.887 ± 0.056	0.847 ± 0.068	0.886 ± 0.039	0.770 ± 0.096	0.928 ± 0.065	0.671 ± 0.086	0.944 ± 0.096	0.460 ± 0.120	1.00 ± 0.000	0.333 ± 0.000
	20%	0.825 ± 0.049	0.876 ± 0.027	0.886 ± 0.103	0.839 ± 0.046	0.912 ± 0.091	0.752 ± 0.061	0.936 ± 0.055	0.728 ± 0.038	1.00 ± 0.000	0.669 ± 0.118	1.00 ± 0.000	0.635 ± 0.062	0.944 ± 0.096	0.511 ± 0.112	0.867 ± 0.231	0.322 ± 0.019
	40%	0.866 ± 0.028	0.928 ± 0.016	0.851 ± 0.034	0.870 ± 0.032	0.974 ± 0.044	0.864 ± 0.022	0.963 ± 0.064	0.666 ± 0.030	1.00 ± 0.000	0.615 ± 0.038	0.952 ± 0.083	0.563 ± 0.015	0.933 ± 0.116	0.385 ± 0.046	1.00 ± 0.000	0.297 ± 0.107
	60%	0.799 ± 0.105	0.886 ± 0.067	0.819 ± 0.091	0.868 ± 0.035	0.877 ± 0.142	0.801 ± 0.040	0.896 ± 0.180	0.718 ± 0.028	0.881 ± 0.206	0.631 ± 0.009	0.939 ± 0.105	0.633 ± 0.061	0.889 ± 0.193	0.447 ± 0.046	0.889 ± 0.193	0.389 ± 0.096
	80%	0.819 ± 0.064	0.890 ± 0.029	0.875 ± 0.008	0.903 ± 0.035	0.905 ± 0.083	0.848 ± 0.035	0.944 ± 0.096	0.711 ± 0.026	0.933 ± 0.116	0.657 ± 0.033	1.00 ± 0.000	0.571 ± 0.000	0.952 ± 0.083	0.463 ± 0.072	1.00 ± 0.000	0.363 ± 0.051
	100%	0.833 ± 0.000	0.909 ± 0.000	0.858 ± 0.030	0.894 ± 0.016	0.923 ± 0.000	0.857 ± 0.000	1.00 ± 0.000	0.767 ± 0.029	1.00 ± 0.000	0.656 ± 0.034	1.00 ± 0.000	0.593 ± 0.038	1.00 ± 0.000	0.524 ± 0.041	1.00 ± 0.000	0.421 ± 0.000
C = 4	5%	0.685 ± 0.052	0.812 ± 0.037	0.696 ± 0.064	0.820 ± 0.046	0.693 ± 0.052	0.785 ± 0.088	0.682 ± 0.067	0.747 ± 0.153	0.665 ± 0.107	0.706 ± 0.214	0.675 ± 0.045	0.630 ± 0.173	0.651 ± 0.073	0.573 ± 0.182	0.565 ± 0.105	0.481 ± 0.129
	10%	0.791 ± 0.042	0.883 ± 0.026	0.812 ± 0.039	0.864 ± 0.032	0.900 ± 0.091	0.859 ± 0.053	0.879 ± 0.036	0.736 ± 0.112	0.958 ± 0.072	0.643 ± 0.094	0.897 ± 0.090	0.531 ± 0.049	1.00 ± 0.000	0.330 ± 0.093	1.00 ± 0.000	0.235 ± 0.000
	20%	0.871 ± 0.050	0.866 ± 0.009	0.972 ± 0.048	0.772 ± 0.037	0.909 ± 0.091	0.719 ± 0.044	0.933 ± 0.116	0.692 ± 0.093	0.958 ± 0.072	0.627 ± 0.016	0.917 ± 0.144	0.439 ± 0.106	0.917 ± 0.144	0.328 ± 0.010	0.889 ± 0.193	0.264 ± 0.061
	40%	0.894 ± 0.034	0.913 ± 0.017	0.910 ± 0.101	0.848 ± 0.048	1.00 ± 0.000	0.732 ± 0.031	1.00 ± 0.000	0.653 ± 0.090	0.952 ± 0.083	0.563 ± 0.015	0.944 ± 0.096	0.466 ± 0.041	1.00 ± 0.000	0.392 ± 0.051	1.00 ± 0.000	0.235 ± 0.000
	60%	0.816 ± 0.063	0.879 ± 0.023	0.907 ± 0.160	0.841 ± 0.051	0.938 ± 0.108	0.743 ± 0.083	0.856 ± 0.171	0.674 ± 0.042	0.909 ± 0.158	0.608 ± 0.033	0.958 ± 0.072	0.560 ± 0.055	0.952 ± 0.083	0.463 ± 0.072	0.889 ± 0.193	0.317 ± 0.074
	80%	0.878 ± 0.004	0.915 ± 0.020	0.903 ± 0.040	0.860 ± 0.035	0.949 ± 0.089	0.744 ± 0.045	0.933 ± 0.116	0.677 ± 0.032	0.958 ± 0.072	0.627 ± 0.016	0.926 ± 0.128	0.552 ± 0.045	0.933 ± 0.116	0.414 ± 0.012	1.00 ± 0.000	0.333 ± 0.000
	100%	0.875 ± 0.000	0.903 ± 0.000	0.906 ± 0.034	0.874 ± 0.021	0.970 ± 0.053	0.756 ± 0.011	1.00 ± 0.000	0.696 ± 0.000	1.00 ± 0.000	0.636 ± 0.000	1.00 ± 0.000	0.571 ± 0.000	1.00 ± 0.000	0.421 ± 0.000	1.00 ± 0.000	0.333 ± 0.000

Observing the results from the perspective of the morphological configurations (rows C), a consistent trend emerges: the optimal performance peak for each row gradually shifts toward lower density values (columns D) as the cleaning intensity increases. This behavior is a direct consequence of the filtering logic of the pipeline. Since morphological cleaning acts by eroding positive pixels, an increase in the parameter C effectively thins out the structural representation of the burrow. Consequently, to maintain a high F1-Score and prevent a collapse in recall, the density threshold D must be proportionally relaxed to accommodate the reduced pixel count within the target area. In contrast, for milder cleaning configurations, a higher number of predicted pixels is maintained, necessitating a stricter density threshold to effectively suppress the surviving background noise. This shift confirms that the two parameters act as a coupled system, where the optimal operating point is defined by their balance rather than by their individual values.

A notable trend is observed in the lower-diagonal region of the parameter matrix, particularly as the training set size increases. In this regime, the system consistently achieves perfect or near-perfect precision. This behavior can be explained by the interplay between model maturity and systemic restrictiveness. As the training set expands from 5% to 100%, the underlying pixel classifier develops a significantly higher discriminative power, producing probability maps with a superior signal-to-noise ratio. When this mature model is paired with aggressive morphological cleaning ($C \geq 3$) and high density thresholds ($D \geq 0.7$), the entry barrier for a candidate detection becomes extremely high. Under these conditions, the probability of background noise or structural artifacts generating a signal strong enough to survive both filters drops to near zero. Consequently, while the system becomes increasingly conservative (at the expense of missing true positive predictions) the reliability of its positive predictions becomes absolute. Such values for C and D can be considered as a high-confidence operating mode, where the total elimination of false alarms is prioritized.

Despite achieving perfect precision in the more restrictive regions (high C and D), the corresponding F1-score reveals a significant performance trade-off. As the system moves toward the lower-diagonal extremity of the grid search, the F1-Score exhibits a noticeable decline, often dropping below the values recorded in the intermediate sweet spot. This divergence is caused by a collapse in recall: the severity of both filters becomes so restrictive that it not only eliminates noise, but also rejects valid burrow candidates. Since the F1-Score balances precision and recall, it penalizes this over-conservative behavior, highlighting that configurations with perfect precision are not necessarily the most effective ones, failing to detect a substantial portion of the targets.

Based on the grid search analysis, the parametrization $C = 2$ and $D = 0.3$ can be identified as the optimal operational configuration. While other specific combinations might yield sporadic performance spikes, this setup consistently delivers state-of-the-art results (achieving a test F1-score of $\approx$0.91 with the full training set) while maintaining superior stability. Crucially, the superiority of this configuration extends beyond the absolute metric values. A defining characteristic of the model obtained by setting $C = 2$ and $D = 0.3$ is its distinct responsiveness to data availability. Unlike minimal configurations (e.g., $C = 0$) which suffered from early saturation and flat performance curves, this setup exhibits a consistent positive learning trend, as seen in Figure 6. As the training set size increases, a noticeable improvement in the metrics is observed, increasing the F1-score from 0.790 (on 5% of the training set) to 0.910 (on 100%). This behavior seems very significant, as it validates the quality of the feature descriptors, proving that the system is not only relying on static filtering constraints, but effectively leverages the additional training examples to refine its decision boundaries and reduce generalization error.

https://doi.org/10.3390/su18042153

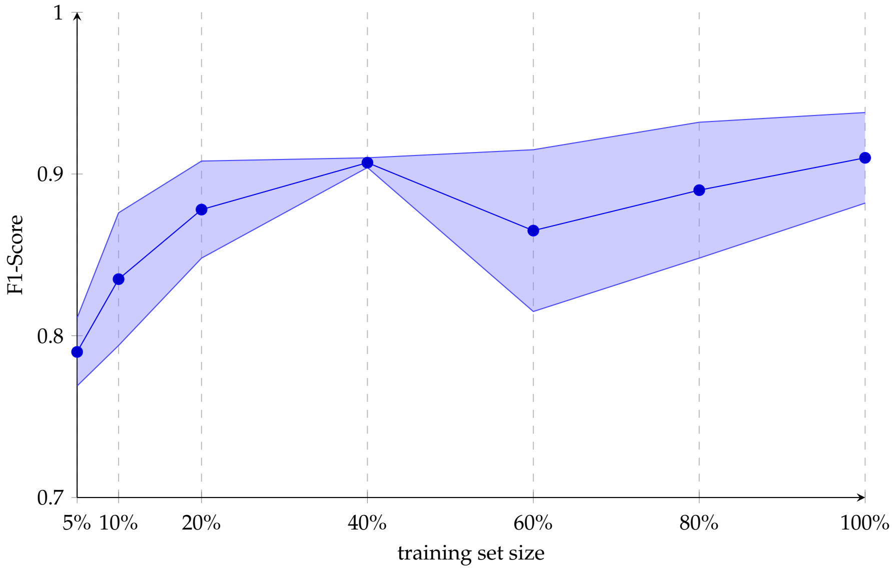


Figure 6. Performance for the best configuration ($C = 2$, $D = 0.3$) varying the size of the training set.

As a final note, it should be observed that our protocol can be immediately extended to cover any similar area; no assumptions have been made on the type of river, riverbanks, or animals involved in the process.

5. Conclusions

This study presents a scalable, data-driven approach for detecting animal burrows in river levee systems using aerial imagery and machine learning techniques. Traditional geophysical methods, while accurate, require specialized equipment, trained personnel, significant time investment and are not always reliable, limiting their applicability for wide monitoring. Our methodology addresses these constraints by leveraging Unmanned Aircraft Systems and Random Forest classification to enable efficient, large scale surveillance of levee infrastructure. The solution addresses both the scarcity and high variability of field data, the latter arising from geographical differences, vegetation density, illumination conditions, image inclination and seasonal factors. Rather than pursuing a single universal model, the system develops a lightweight customized model, tailored to the collected images belonging to a specific monitoring campaign. The classification workflow addresses a fundamental challenge in infrastructure monitoring: the need to inspect extensive levee networks efficiently. Rather than requiring users to manually examine every image in detail, the system enables them to label only a small number of representative regions that exemplify burrow and non-burrow characteristics. From these limited training examples, a Random Forest classifier learns to automatically process entire image datasets, identifying potential burrow locations across hundreds or thousands of images. The automated detection pipeline applies morphological filtering, density analysis, and geometric constraints to refine the initial classifications, presenting users with a curated set of candidate locations for final validation.

This process went through comprehensive experimental validation using a full grid search strategy, exploring the interplay between cleaning configurations and spatial density thresholds across varying training set sizes. Results revealed that a compensatory relationship existed between the parameters of the system; the optimal parametrization allowed us to achieve a good balance between precision and recall even with limited training data, with a performance stably between 0.8 and 0.9 F1-score, effectively suggesting that our approach may be feasible. This last result is particularly valuable for operational deployment, as it provides users with an economical annotation workflow, requiring minimal manual labeling effort to achieve reliable detection across large image datasets. This dramatically reduces the inspection burden, transforming a task that would require hours of careful manual review into a process where human expertise is focused only on providing initial guidance and validating final results. This research demonstrates the viability of automation and machine learning for infrastructure monitoring, offering a practical alternative, as well as a supporting tool, to traditional survey methods. By providing precise threat identification and efficient monitoring capabilities, the system facilitates a fundamental shift from reactive emergency repairs to proactive maintenance strategies.

The software has already been adopted by public officials for operational use, demonstrating its practical value in real-world monitoring scenarios. Such practical deployment establishes a continuous feedback loop between field operators and system development, ensuring that it evolves in response to actual monitoring challenges. Furthermore, it provides operational data collected through routine use, creating opportunities for expanding the training dataset across diverse geographical contexts, progressively enhancing the model's generalization capabilities.

Author Contributions: Conceptualization, A.C. and F.P.; methodology, G.S.; software, G.P. and E.P.; validation, L.B., M.G. and M.V.; formal analysis, G.S.; investigation, A.P. and S.P.; resources, R.Z. and D.M.; data curation, D.M.; writing—original draft preparation, G.P. and E.P.; writing—review and editing, G.S. and A.P.; visualization, G.P.; supervision, R.Z. and G.S.; project administration, R.Z.; funding acquisition, R.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received was funded by Emilia-Romagna region.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Our original dataset is available at https://www.kaggle.com/dsv/14153346 (accessed on 15 November 2025), with DOI 10.34740/KAGGLE/DSV/14153346.

Conflicts of Interest: Davide Martinucci and Roberta Zambrini were employed by Esplora S.R.L. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Federal Emergency Management Agency (FEMA). *Technical Manual for Dam Owners: Impacts of Animals on Earthen Dams*; Technical Manual no. 473; Federal Emergency Management Agency (FEMA): Washington, DC, USA, 2005.
2. Sharp, M.; Wallis, M.; Deniaud, F.; Hersch-Burdick, R.; Tourment, R.; Matheu, E.; Seda-Sanabria, Y.; Wersching, S.; Veylon, G.; Durand, E.; et al. *The International Levee Handbook*; CIRIA: London, UK, 2013.
3. Bayoumi, A.; Meguid, M.A. Wildlife and safety of earthen structures: A review. *J. Fail. Anal. Prev.* **2011**, *11*, 295–319. [CrossRef]
4. Richards, K.S.; Reddy, K.R. True triaxial piping test apparatus for evaluation of piping potential in earth structures. *Geotech. Test. J.* **2010**, *33*, 83–95. [CrossRef]
5. Orlandini, S.; Moretti, G.; Albertson, J.D. Evidence of an emerging levee failure mechanism causing disastrous floods in Italy. *Water Resour. Res.* **2015**, *51*, 7995–8011. [CrossRef]
6. Camici, S.; Barbetta, S.; Moramarco, T. Levee body vulnerability to seepage: The case study of the levee failure along the Foenna stream on 1 January 2006 (Central Italy). *J. Flood Risk Manag.* **2017**, *10*, 314–325. [CrossRef]
7. Borgatti, L.; Forte, E.; Mocnik, A.; Zambrini, R.; Cervi, F.; Martinucci, D.; Pellegrini, F.; Pillon, S.; Prizzon, A.; Zamariolo, A. Detection and characterization of animal burrows within river embankments by means of coupled remote sensing and geophysical techniques: Lessons from River Panaro (northern Italy). *Eng. Geol.* **2017**, *226*, 277–289. [CrossRef]
8. Palladino, M.R.; Barbetta, S.; Camici, S.; Claps, P.; Moramarco, T. Impact of animal burrows on earthen levee body vulnerability to seepage. *J. Flood Risk Manag.* **2019**, *13*, e12559. [CrossRef]
9. Panthulu, T.; Krishnaiah, C.; Shirke, J. Detection of seepage paths in earth dams using self-potential and electrical resistivity methods. *Eng. Geol.* **2001**, *59*, 281–295. [CrossRef]
10. Butler, J.; Roper, T.J.; Clark, A.J. Investigation of badger (*Meles meles*) setts using soil resistivity measurements. *J. Zool.* **1994**, *232*, 409–418. [CrossRef]

11. Ni, S.H.; Chen, C.K.; Lin, H.M. Application of Ground Penetrating Radar on the Void-Detection in Levee. In *Proceedings of the ISOPE International Ocean and Polar Engineering Conference, Kitakyushu, Japan, 26–31 May 2002*; ISOPE: Cupertino, CA, USA, 2002; pp. 836–839.
12. Pennisi, V.; Musumeci, R.E.; Grasso, R.; Ferraro, A.; Foti, E. Non-invasive surveys to assess animal burrows as potential causes of levee instability. *J. Flood Risk Manag.* **2023**, *16*, e12931. [CrossRef]
13. Loperte, A.; Bavusi, M.; Cerverizzo, G.; Lapenna, V.; Soldovieri, F. Ground penetrating radar in dam monitoring: The test case of Acerenza (Southern Italy). *Int. J. Geophys.* **2011**, *2011*, 654194. [CrossRef]
14. Bertone, N.; Forte, E.; Titti, G.; Zambrini, R.; Macini, P.; Mocnik, A.; Parodi, S.; Pellegrini, F.; Possamai, M.; Staboli, C.; et al. Integration of seasonal frequency domain electromagnetic surveys and geological data for assessing the integrity of earthen levee systems. The case study of the Panaro River (Northern Italy). *Eng. Geol.* **2024**, *342*, 107749. [CrossRef]
15. Anderson, N.; Ismail, A. A generalized protocol for selecting appropriate geophysical techniques. In *Proceedings of the Geophysical Technologies for Detecting Underground Coal Mine Voids Forum, Lexington, KY, USA, 28–30 July 2003*; Citeseer: University Park, PA, USA, 2003; pp. 28–30.
16. Benson, R.C. Remote sensing and geophysical methods for evaluation of subsurface conditions. In *Practical Handbook of Environmental Site Characterization and Ground-Water Monitoring*; CRC Press: Boca Raton, FL, USA, 2005; pp. 259–306.
17. Butler, D.K. Microgravimetric and gravity gradient techniques for detection of subsurface cavities. *Geophysics* **1984**, *49*, 1084–1096. [CrossRef]
18. Roth, M.; Mackey, J.; Mackey, C.; Nyquist, J. A case study of the reliability of multielectrode earth resistivity testing for geotechnical investigations in karst terrains. *Eng. Geol.* **2002**, *65*, 225–232. [CrossRef]
19. Szynkiewicz, A. GPR monitoring of earthen flood banks/levees. In *Proceedings of the Eighth International Conference on Ground Penetrating Radar, Gold Coast, Australia, 23–26 May 2000*; SPIE: Bellingham, WA, USA, 2000; Volume 4084, pp. 85–90.
20. Kinlaw, A.; Grasmueck, M. Evidence for and geomorphologic consequences of a reptilian ecosystem engineer: The burrowing cascade initiated by the gopher tortoise. *Geomorphology* **2012**, *157*, 108–121. [CrossRef]
21. Cortez, J.D.; Henke, S.E.; Redeker, E.; Fulbright, T.E.; Riddle, R.; Young, J. Demonstration of ground-penetrating radar as a useful tool for assessing pocket gopher burrows. *Wildl. Soc. Bull.* **2013**, *37*, 428–432. [CrossRef]
22. Di Prinzio, M.; Bittelli, M.; Castellarin, A.; Pisa, P.R. Application of GPR to the monitoring of river embankments. *J. Appl. Geophys.* **2010**, *71*, 53–61. [CrossRef]
23. van Lieshout, C.; van Oeveren, K.; van Emmerik, T.; Postma, E. Automated River Plastic Monitoring Using Deep Learning and Cameras. *Earth Space Sci.* **2020**, *7*, e2019EA000960. [CrossRef]
24. Hidaka, M.; Matsuoka, D.; Sugiyama, D.; Murakami, K.; Kako, S. Pixel-level image classification for detecting beach litter using a deep learning approach. *Mar. Pollut. Bull.* **2022**, *175*, 113371. [CrossRef]
25. Gonçalves, G.; Andriolo, U.; Pinto, L.; Duarte, D. Mapping marine litter with Unmanned Aerial Systems: A showcase comparison among manual image screening and machine learning techniques. *Mar. Pollut. Bull.* **2020**, *155*, 111158. [CrossRef]
26. Kako, S.; Morita, S.; Taneda, T. Estimation of plastic marine debris volumes on beaches using unmanned aerial vehicles and image processing based on deep learning. *Mar. Pollut. Bull.* **2020**, *155*, 111127. [CrossRef]
27. Bao, Z.; Sha, J.; Li, X.; Hanchiso, T.; Shifaw, E. Monitoring of beach litter by automatic interpretation of unmanned aerial vehicle images using the segmentation threshold method. *Mar. Pollut. Bull.* **2018**, *137*, 388–398. [CrossRef] [PubMed]
28. Deidun, A.; Gauci, A.; Lagorio, S.; Galgani, F. Optimising beached litter monitoring protocols through aerial imagery. *Mar. Pollut. Bull.* **2018**, *131*, 212–217. [CrossRef] [PubMed]
29. Fallati, L.; Polidori, A.; Salvatore, C.; Saponari, L.; Savini, A.; Galli, P. Anthropogenic Marine Debris assessment with Unmanned Aerial Vehicle imagery and deep learning: A case study along the beaches of the Republic of Maldives. *Sci. Total Environ.* **2019**, *693*, 133581. [CrossRef] [PubMed]
30. Moy, K.; Neilson, B.; Chung, A.; Meadows, A.; Castrence, M.; Ambagis, S.; Davidson, K. Mapping coastal marine debris using aerial imagery and spatial analysis. *Mar. Pollut. Bull.* **2018**, *132*, 52–59. [CrossRef]
31. Nayak, R.; Pati, U.C.; Das, S.K. A comprehensive review on deep learning-based methods for video anomaly detection. *Image Vis. Comput.* **2021**, *106*, 104078. [CrossRef]
32. Sifnaios, S.; Arvanitakis, G.; Konstantinidis, F.K.; Tsimiklis, G.; Amditis, A.; Frangos, P. A deep learning approach for pixel-level material classification via hyperspectral imaging. *arXiv* **2024**, arXiv:2409.13498.
33. Ramesh, K.; Kumar, G.K.; Swapna, K.; Datta, D.; Rajest, S.S. A review of medical image segmentation algorithms. *EAI Endorsed Trans. Pervasive Health Technol.* **2021**, *7*, e1–e13. [CrossRef]
34. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
35. Gorishniy, Y.; Rubachev, I.; Khrulkov, V.; Babenko, A. Revisiting deep learning models for tabular data. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 18932–18943.
36. Shwartz-Ziv, R.; Armon, A. Tabular data: Deep learning is not all you need. *Inf. Fusion* **2022**, *81*, 84–90. [CrossRef]

37. Grinsztajn, L.; Oyallon, E.; Varoquaux, G. Why do tree-based models still outperform deep learning on typical tabular data? *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 507–520.
38. Baramal, S.; Chauhan, D. *A Comparative Study of TabNet and XGBoost for Tabular Data Classification*; Springer Nature: Berlin, Germany, 2025.
39. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
40. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [CrossRef]
41. Sheykhmousa, M.; Mahdianpari, M.; Ghanbari, H.; Mohammadimanesh, F.; Ghamisi, P.; Homayouni, S. Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6308–6325. [CrossRef]
42. Feng, Q.; Liu, J.; Gong, J. UAV remote sensing for urban vegetation mapping using random forest and texture analysis. *Remote Sens.* **2015**, *7*, 1074–1094. [CrossRef]
43. Grimmer, G.; Wenger, R.; Forestier, G.; Chardon, V. Automatic detection of in-stream river wood from random forest machine learning and exogenous indices using very high-resolution aerial imagery. *Environ. Model. Softw.* **2025**, *190*, 106460. [CrossRef]