DOI 10.13131/UNIPI/RW9Q-B777

Caterina Paola Venditti*, Antonio Leopardi**

* University of Cassino and Southern Lazio – European University of Technology.

** IACnet.

PROCESSING TOPOGRAPHICAL MAPS FOR ARCHAEOLOGICAL RESEARCH: FROM A "ZERO-CODE" TO A "LOW-CODE" AI APPROACH

DATASET: DOI 10.13131/UNIPI/ZMM8-0K39

Abstract: The analysis of cartographic sources is fundamental for reconstructing ancient landscapes and understanding long-term territorial transformations. Archaeological maps produced between the early twentieth century and the advent of digital cartography preserve valuable spatial and contextual information, yet their manual transcription remains time-consuming and labour-intensive. This study proposes an AI-based workflow for the semi-automated recognition, extraction, and georeferencing of cartographic features from analogue maps. The workflow integrates free and low-cost tools for object detection and segmentation, combining computer vision models with automated support for data annotation and processing. Both map-level and symbol-level strategies are explored to address cartographic complexity and iconographic variability. While challenges remain in adapting models to specialised symbols and ensuring spatial accuracy, the results demonstrate the effectiveness of the approach, with 272 automatically detected features. The highest performance metrics on test set are obtained for the most represented class in the training dataset (a_f: Box mAP50 = 0.981; Mask mAP50-95 = 0.606), with average values across all classes reaching a Box mAP50 = 0.979 and a Mask mAP50-95 = 0.573. Furthermore, they confirm the feasibility of converting historical maps into vector datasets suitable for integration into GIS environments.
Keywords: historical maps, deep learning, feature extraction, GIS integration.

1. Introduction

The recording and transcription of historical cartographic data, from hydrographic and orographic components to the diverse manifestations of human presence, have long constituted a pivotal step in constructing interpretative frameworks capable of clarifying the dynamics of landscape transformation[1]. Such practices have always been

[1] The use of historical cartography as a primary source for reconstructing ancient landscapes is deeply rooted across all major schools of topography, and the Italian tradition has played a particularly significant role in shaping this perspective. These scholarly lineages established the integration of maps, field observation,

crucial, and at times even urgent for purposes beyond pure research, traditionally relying on slow and attentive processes in which the very act of reading maps contributed directly to understanding landscape history. Today, the possibilities opened up by machine learning define a new horizon in the acquisition and management of information from maps, by accelerating and refining processes of data extraction and the construction and normalisation of derived geospatial datasets[2].

It is precisely in this context that AI4MAPS was conceived as an exploratory research initiative aimed at investigating AI-assisted solutions for the extraction and spatial integration of archaeological data from cartography[3]. More specifically, this paper investigates how archaeological features can be automatically recognised and extracted from raster maps, and how these outputs can be transformed into GIS-ready spatial data suitable for further analysis.

The research is guided by two main questions: (1) to what extent can computer vision models reliably detect and segment archaeological symbols embedded within complex historical cartographic backgrounds? and (2) how can the extracted information be efficiently converted into interoperable geospatial datasets for GIS incorporation and landscape analysis?

To address these questions, the study develops and evaluates a lightweight and modular workflow based on freely available tools (combining object detection, segmentation, symbol-level classification) and custom Python-based pipelines.

2. State of the Art, Project Rationale, and Challenges

Previous research

The study and optimisation of systems for automated analysis of cartography in its broad sense has already generated a substantial amount of research[4], and recent literature reviews have highlighted an increasing engagement with AI methods and several major research lines.

---

and historical interpretation as a core methodological foundation for archaeological research, a framework that continues to inform contemporary analyses.

[2] In recent years, the number of studies applying ML to historical cartography has grown exponentially; see, e.g., García-Molsosa *et al.*, 2021; Göderle *et al.*, 2023; Fylaktos *et al.*, 2025, with large references.

[3] The present study is rooted in a line of research initiated approximately at the end of 2024, originally conceived as an exploratory and provocative attempt to assess whether a fully LLM-based approach could be used for the automatic recognition of symbols on archaeological maps, without relying on complex platforms or dedicated code (Venditti, 2025). Although these preliminary experiments proved valuable in assessing the potential of AI for the selected tasks, they also exposed the fuzzy nature of multimodal inference in general-purpose language models for object detection and feature extraction, underscoring the need for more specialised tools and dedicated computational environments.

[4] This is a rapidly evolving field, in which it is increasingly difficult to track parallel developments pursued by different research groups with convergent, though not necessarily aligned, objectives. An interesting literature review on various applications of AI in cartography can be found in Kang *et al.*, 2024 and Mai *et al.*, 2025.

One particularly prominent direction focuses on the semantic exploration of large historical map collections, exemplified by major projects such as MapReader, which has developed tools used to automate the detection of infrastructural features, including railways and other linear elements (Hosseini *et al.*, 2022)[5]. Complementing these efforts, other studies target the extraction and georeferencing of textual information, most notably mapKurator, an end-to-end system that automates OCR, post-processing, and the spatial linking of text labels from large-format map scans (Kim *et al.*, 2023)[6].

More recent LLM- and attention-based methods have demonstrated strong performance in annotating cartographic content and detecting symbols. Studies combining YOLO with attention modules have achieved high precision and recall in recognising features on complex images (Jiang *et al.*, 2022, for general methodological efficacy; Liu *et al.*, 2023, for complex background applications). Comparable results were obtained by an enhanced YOLOv3 architecture integrating Convolutional Block Attention Module (CBAM) for the detection of urban Points of Interest (POIs) on georeferenced digital maps, with detected symbols automatically linked to corresponding features in the underlying database (Zhang *et al.*, 2022)[7]. Based on similar architectures, other scholars focus on the automatic extraction of cartographic features (Xydas *et al.*, 2022; Jiao *et al.*, 2022; Martinez *et al.*, 2023; O'Hara *et al.*, 2024; Saxton *et al.*, 2024; Kramm *et al.*, 2025; McCarthy, 2025), such as buildings, road networks, and selected natural environments from maps like land-cover classes, while ignoring surrounding or overlapping information.

Moving closer to the archaeological domain, research on early twentieth-century French colonial cartography applied a segmentation-based method with multiple U-Net detectors trained on georeferenced maps to identify archaeologically meaningful symbols such as the "*ruins*" marker (García-Molsosa *et al.*, 2021)[8]. Similarly, the RePaSE project developed a browser-based tool using DeepLabv3 to extract building footprints from georeferenced cadastral maps, aerial photographs, and satellite imagery with high accuracy (Göderle *et al.*, 2023; Göderle *et al.*, 2024). From the standpoint of georeferencing, a broader methodological perspective is offered by the "*Artificial Intelligence for Critical Mineral Resource Assessment*" initiative (Goldman *et al.*, 2025), which demonstrated how the automation of georeferencing can substantially reduce the time and labour traditionally required to convert legacy raster maps into analysis-ready datasets. In addition, these studies intersect with the objectives of the ICA's Commission on Cartographic Heritage into the Digital, which advocates the

[5] See also Wood *et al.*, 2024, with references.

[6] See also https://knowledge-computing.github.io/mapkurator-doc/#/docs/introduction (retrieved December 4th, 2025).

[7] See also the pipeline proposed in Cao *et al.*, 2025 for icon automated extraction combining YOLO and SAM.

[8] The authors implemented the models using the Picterra platform (https://picterra.ai/ retrieved December 4th, 2025), which automates data augmentation, cloud-based training, and post-processing of instance segmentation outputs.

integration of advanced digital and AI-based methods to enhance the georeferencing, preservation, and spatial analysis processes[9].

Recent initiatives outside the academic domain further confirm the growing relevance of AI-based approaches for deriving structured information from cartographic sources. A notable example is the UK government project *Extract*[10], developed to unlock data from legacy planning documents by automatically parsing textual and spatial information and converting it into GIS-ready formats. Although *Extract* is designed for urban planning rather than archaeological analysis, it demonstrates how similar challenges are currently being addressed at an institutional level and highlights the broader applicability of AI in this domain.

Positioning AI4MAPS project

Within the broader scenario outlined above, archaeological cartography remains underrepresented. Addressing this emerging research area, the present project adopts a discipline-specific perspective tailored to the representational conventions of archaeological maps. The two main research questions introduced earlier define a framework for practical, AI-assisted extraction and GIS integration of archaeological cartographic data.

The overarching aim is to enable large-scale, rapid retrieval of information from extensive map *corpora*, thereby supporting timely analyses of landscape changes. In this sense, the project contributes to ongoing efforts toward increasingly advanced forms of automated map interpretation, potentially opening pathways to future applications such as automated derivation of DEMs from historical cartographic sources.

In contrast to resource-intensive pipelines, the project evaluates lightweight, modular workflows, compatible with freely available tools, limiting reliance on complex coding or specialist computational infrastructures. The aim is to enable researchers and heritage practitioners to digitise and spatially integrate archaeological cartography even in low-resource settings and, in doing so, to contribute to a research area that remains comparatively underexplored.

Main Challenges

The challenges involved span multiple dimensions. Some stem from the quality, consistency, and resolution of the underlying digital data, while others concern model performance, including the generation of GIS-ready features, the handling of segmentation complexity, and the need to generalise without overfitting[11].

A central source of difficulty, however, lies in the intrinsic characteristics of archaeological cartography, such as symbol variability, heterogeneous base maps, and graphic

[9] The annual international conference "Digital Approaches to Cartographic Heritage" serves as a major forum for presenting advances in the digital treatment and analysis of historical maps (https://cartography.web.auth.gr/ICA-Heritage/index.html, retrieved December 4th, 2025).

[10] About the project see https://extract.ai.gov.uk/ (retrieved December 10th, 2025).

[11] Useful strategies to mitigate overfitting in Ying, 2019.

ambiguity[12]. This heterogeneity largely derives from fluctuations in the graphic language, reflecting not simply technical choices but also culturally embedded practices of representation. Traditional conventions placed raw data (e.g., *areas with pottery sherds, masonry structures*) and archaeological interpretation (e.g., *villa, tower, necropolis*) on the same semiotic level, encoding as graphic symbols both observed features and the interpretative hypotheses built upon them[13]. While such informational density was meaningful within analogue mapping practices, it generated a proliferation of variables, some corresponding to differences in meaning, others to differences in signifiers. In digital environments, where content is managed as discrete logical-numerical entities, this ambiguity becomes more problematic and requires greater analytical precision in separating empirical evidence from interpretative constructs.

3. Data

The present work focuses on a particular type of archaeological cartography, namely the *Forma Italiae* series[14] and *Atlante Tematico di Topografia Antica* series[15], because, for the characteristics outlined below, they constitute a particularly valuable *corpus* for this study. Over time, both series have undergone significant changes, including variations in colour schemes, information density, and the range of geodata represented. Archaeological content is clearly distinguished from the topographic background, often through the use of distinctive colour variables that enhance visual separation, while the symbol set employed is heterogeneous, comprising point, polygonal, and linear features, with various graphical rendering, typically accompanied by alphanumeric identifiers that support systematic cataloguing.

Substantial efforts are currently underway to digitise and organise the data recorded on these maps, as exemplified by initiatives such as the *TIR-FOR Project*, which involve the *Forma Italiae* project (Migliorati & Leopardi, in press)[16]. The

[12] Indeed, one can observe a degree of inconsistency in the selection of characteristics or attributes used to derive the graphical variables for representation, constituting an interesting testing ground for a broader extraction of graphic elements from analogue cartographic sources (*infra*).

[13] This oscillation is particularly evident in long-standing cartographic initiatives, such as the *Forma Italiae* project, up to the transition toward the full digital management of data.

[14] The *Forma Italiae* project, founded by Giuseppe Lugli in 1923, was established under the auspices of the Accademia dei Lincei and the National Academic Union with the intention of producing an Archaeological Map of Italy. The project continues today under the direction of Luisa Migliorati and with the patronage of the Union Académique Internationale. For a recent overview of the Italian cartographic tradition and earlier bibliography, see Marchi, 2023.

[15] The *Atlante Tematico di Topografia Antica* (ATTA) was established in 1990 and brings together the work of numerous scholars from universities, civil servants from the Ministry of Culture, various research institutions, public and private organisations, as well as independent researchers. Specific topics are developed in the *Supplementi* of the journal, including the topographical and archaeological investigations conducted in Campania since 1999 under the auspices of the Chair of Ancient Topography at the Second University of Naples.

[16] https://tir-for.iec.cat/ (retrieved December 5th, 2025).

progressive availability of such digital resources holds considerable potential for building more robust training datasets and improving the performance of automated feature-retrieval.

At the same time, the richness of the topographic base introduces an additional layer of complexity. Hydrography, vegetation cover, contour lines, and a range of other environmental and anthropic elements, with their value for landscape interpretation, also visually intertwined with the underlying cartographic detail. From this perspective, AI4MAPS also aims to extend the method to the extraction of selected landscape features, supporting more comprehensive landscape-archaeological analysis.

The dataset used in this study is based on selected sheets drawn at a scale of 1:25,000, which provides sufficient detail to document both settlement structures and features of the surrounding landscape[17]. A known limitation derives from the nature of the available material, which includes maps reproduced through both photographic capture and direct scanning. As a result, the images inevitably present some distortions, variations in colour balance, and differences in resolution that reflect the mixed-nature of the original sources[18]. However, such images offer a realistic testing ground for recognition approaches, as they mirror the imperfect and diverse nature of cartographic materials commonly encountered in research.

A further limitation concerns the size of the initial dataset, which was inevitably small due to the availability of cartographic material at this early, exploratory stage of the project. For training, 132 images corresponding to map excerpts were used, with a total of 1,872 annotations distributed across three classes representing three different symbols (two point-based and one area-based). Though the sample size, the aim at this stage was primarily to test the workflow and assess its feasibility. Data augmentation was applied during training to increase dataset variability and improve model generalisation.

4. Methodology

When considering feature-extraction tasks, traditional methods focused on low-level visual properties such as edges, colours, or textures, but they were highly sensitive to noise, resolution, and stylistic variation[19]. More recently, object-detection and

[17] The dataset derives from a published cartographic series and, as such, they are publicly accessible through libraries and research institutions. The scans and photographs used in this study were generated for research purposes; while the printed volumes are under copyright, their use here falls within scholarly and non-commercial contexts, with due bibliographic reference to the original publication (D. Lgs. 177/2021, Directive (EU) 2019/790).

[18] Image distortions and acquisition-related artefacts are known to affect the robustness and accuracy of convolutional neural networks (Hendrycks & Dietterich, 2019).

[19] The pixel-based trend derives directly from classic image processing (Gonzalez & Woods, 1992; see also Chiang *et al.*, 2014). New lines of research in this path involve the integration of Object-Based Image Analysis techniques and pre-existing text detection and OCR tools, suitably adapted to the context of historical cartographic sources (Schlegel, 2021; Schlegel, 2023).

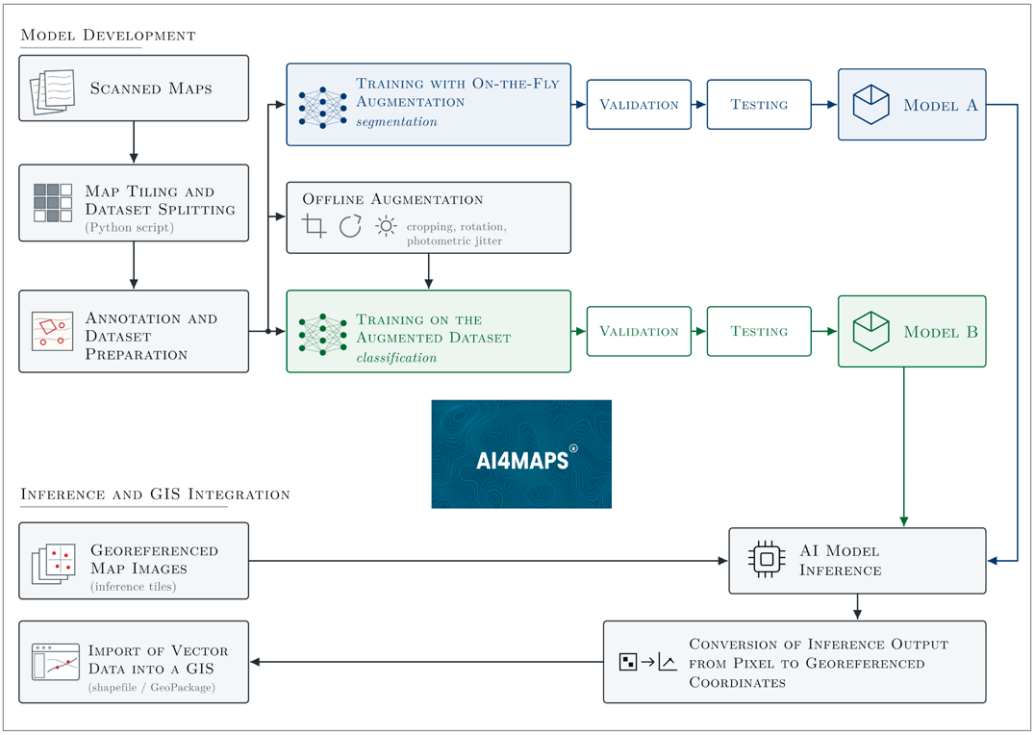

fig. 1. AI4MAPS Workflow. Model A is trained as an instance segmentation model to detect individual cartographic symbols and predict a bounding box and a segmentation mask for each instance. The annotated symbols are used to generate an augmented classification dataset for training Model B, which assigns each symbol to its corresponding cartographic class.

segmentation frameworks have introduced an object-level perspective[20], enabling the more robust recognition of cartographic symbols as discrete graphical entities. The AI4MAPS project adopts the latter approach by implementing a fully customisable environment in Google Colab, chosen for its rapid deployment and free access to GPU resources, combining conversational models for code generation with transfer learning techniques applied to pretrained YOLO models (**fig. 1**). The proposed workflow comprised the development of two models. The first (Model A) was designed for the automatic segmentation of archaeological symbols on historical maps; the annotated dataset created for training the segmentation model was subsequently reused to generate the training dataset for the second model (Model B), which classified the symbols into the predefined archaeological categories. While the environment proved useful for initial parameter calibration, several structural limitations became evident during the work sessions (time-limited runtime sessions, inconsistency in the type

[20] These methods recognise symbols as discrete entities rather than pixel patterns, learning higher-level visual features that are more robust to variation, scale, noise, and distortion. Their outputs integrate smoothly with vector-based GIS workflows, providing a more scalable and effective approach to cartographic feature extraction.

and performance of allocated GPUs, and the need for systematic manual saving of models and logs to ensure reproducibility)[21]. Consequently, it became clear that this solution is less suitable for prolonged training phases or large-scale hyperparameter experimentation.

Selecting the model

The choice of software and tools aligns with the overarching goal of adopting an accessible and reproducible methodology based on free and widely available resources, ensuring that the entire procedure can be applied even outside highly specialised computational environments or technical expertise. Model A training was carried out using the pretrained YOLOv8n-seg architecture provided by the Ultralytics framework, allowing the network to adapt its general visual features to domain-specific cartographic symbols through transfer learning[22]. This architecture is particularly suitable for the task because segmentation networks can identify and localise multiple symbols directly within full map sheets, while also delineating their precise shape through pixel-level masks. This enables the system to handle complex layouts, overlapping graphical elements, and heterogeneous cartographic styles with greater expressive power than bounding box-based ones. When configured with appropriate data augmentation and higher image resolutions (e.g., *imgsz = 1024* or above), YOLOv8 performs well effectively even on small graphical objects, an essential requirement for symbol detection on historical maps. Additional advantages include extensive online documentation, an active user community, and a range of model sizes adaptable to different computational constraints, including *nano* and *small* variants suited to rapid experimentation, and *medium* or *large* versions offering higher accuracy.

Data Preparation

During pre-processing, images were subdivided into tiles using a Python script to generate the annotation dataset. To ensure consistent input during training, all maps were standardised to a uniform size of 1280×1280 pixels, with a 15% overlap between adjacent tiles to reduce the risk of features being truncated at the edges.

Images were annotated using CVAT, a free web-based platform[23], and archaeological symbols were manually labelled according to three recurrent classes observed in historical survey maps:

---

[21] Sessions are time-limited (typically 3-12 hours in the free tier) and may terminate due to inactivity or high usage; memory is constrained (approximately 12-25 GB) and local storage is temporary, being cleared at the end of each session; administrative privileges are unavailable, and intensive GPU use may temporarily reduce access to these resources.

[22] YOLOv8 was chosen because at the time of testing it was the most stable version of the YOLO family: it provides high precision in detecting fine graphic details, making it suitable for the present application, but lacks support for image georeferencing metadata.

[23] CVAT (Computer Vision Annotation Tool) was chosen as it allows users to upload batches of images and define annotation tasks with custom object classes, and it is available as a free, open-source platform (https://www.cvat.ai/, retrieved August 10th, 2025).

– Class 0 (*elemento_puntuale* = *e_p*) - single solid dot, representing punctual archaeological element (e.g., small sites or findspots), annotated with tight ellipses or small polygonal masks (599 original instances + augmented samples generated through data augmentation);
– Class 1 (*area_frammenti* = *a_f*) - dense group or cluster of dots, corresponding to areas of scattered ceramic fragments, annotated with free-form polygons to capture their irregular outlines (1076 original instances + augmented samples generated through data augmentation);
– Class 2 (*materiale_sporadico* = *m_s*) - three dots forming a triangle, indicating sporadic surface finds, annotated using triangular polygonal masks (85 original instances + augmented samples generated through data augmentation).

These three classes share a common visual characteristic as they are composed of one or more red dots. This deliberate choice was intended to increase the complexity of the detection task, and to intentionally challenge the model by introducing high inter-class similarity, particularly in the case of fragment-scatter areas, whose geometry is irregular and never identical, while testing its ability to discriminate subtle morphological differences. The annotations were then exported in YOLO segmentation format, generating a separate *.txt* file for each image containing the class identifier followed by the pixel normalised coordinates of all polygon vertices describing each symbol's mask.

A Python script was also used to split the dataset into training, validation, and test subsets (70% / 20% / 10%), ensuring a balanced distribution of classes. The segmentation labels exported in YOLOv8 format were then matched to their corresponding images[24].

The training dataset underwent a targeted manual and automated cleaning and validation process. The built-in verification mechanisms of CVAT helped constrain several common geometric and topological annotation issues, including edge misalignment, polygon redundancy, and label inconsistencies. Additional Python-based checks were then used to verify file consistency and the correspondence between images and annotation files. Manual cleaning was instead limited to the visual inspection of the generated image tiles: this step was primarily aimed at identifying and removing outlier tiles containing extensive blank or unmapped regions, such as margins or scanner border artefacts, which would otherwise introduce uninformative gradients during training[25].

No heavy pre-processing was intentionally applied, given the specific nature of the historical cartographic material represented in the images. Indeed, these documents contain high-frequency visual details and very fine linear features that are essential for the recognition of archaeological symbols. Standard denoising procedures, such as Gaussian or bilateral smoothing, risk blurring these boundaries and degrading the structural information required for accurate segmentation.

---

[24] A separate dataset of symbol crops was also generated from the CVAT annotations for the purpose of training an independent classification model (Model B). See *infra* par. 7.

[25] A small number of tiles (3.7% of the initial dataset), corresponding either to nearly blank map portions or to cartographic content not relevant to the selected analytical classes, were detected and excluded prior to training.

Model Training & Evaluation

The training configuration was designed to balance computational efficiency with the need for adequate model convergence. Training was performed for 50 epochs with a batch size of 4, a compromise chosen to ensure stable convergence while accommodating GPU memory constraints in the segmentation task. An explicit early-stopping criterion was enabled with patience of 20, allowing the effective number of epochs to adapt to the actual learning process, potentially reducing training time and further mitigating overfitting[26].

Given the relatively small number of maps, whole-map augmentation was applied on-the-fly during training to increase data diversity without physically enlarging the dataset. The applied transformations were deliberately light and context-preserving, introducing controlled variability in map appearance (e.g., printing artefacts, lighting conditions, and orientation) while maintaining realistic cartographic layouts. These included rotations of ±5°, minor translations (≤5% of image size), light scaling (±10%), minimal shear (2.0), horizontal flipping[27] with a probability of 0.5, and a very small perspective change (0.0005).

Model training was conducted using the YOLOv8n-seg architecture on a Tesla T4 GPU (15 GB VRAM) allocated through Google Colab (free tier). The standard architecture was employed without introducing additional custom layers or structural modifications, and the pretrained weights were fine-tuned on the custom dataset. The full 50-epoch training procedure was completed in approximately 0.176 hours (~10.5 minutes), confirming the efficiency of the selected architecture in lightweight cloud-based environments[28]. The best-performing checkpoint was selected automatically during training based on validation performance, and all training outputs (model weights, logs, and evaluation metrics) were stored in Colab's directory.

Validation Results

During training, model optimization is monitored through the training loss components (Box Loss, Segmentation Loss, Classification Loss and Distribution Focal Loss, DFL), which are reported in **tab. 1a**. Evaluation on the validation set (32 images, 533 annotated instances) demonstrates consistently strong performance across the classes. To facilitate direct comparison between training, validation and held-out test performance, the same evaluation metrics were computed using the standard YOLOv8 evaluation procedure (**tab. 1b**)[29].

---

[26] No exhaustive hyperparameter optimisation was conducted at this exploratory stage of the research. The final training configuration was selected through preliminary empirical testing, balancing validation stability, computational cost, and GPU memory constraints.

[27] Conversely, vertical flipping and large rotations were deliberately disabled to avoid unrealistic map orientations.

[28] According to the training logs, the resulting YOLOv8n-seg model comprised 85 layers, 3.26 million parameters, and approximately 11.3 GFLOPs.

[29] In YOLO-based models, Precision (P), Recall (R), mAP@50 and mAP@50-95 are computed on the validation and test datasets rather than on the training set. This evaluation protocol prevents overly

Metrics	Train box loss	Train segmentation loss	Train classification loss	Train DFL loss
Final value (Epoch 50)	1.415	1.871	1.097	1.013

tab. 1a. Final training loss metrics after 50 epochs.

Split	Class	Instances	Box P	Box R	Box mAP50	Box mAP50-95	Mask P	Mask R	Mask mAP50	Mask mAP50-95
Train	all	1630	0.926	0.813	0.876	0.549	0.925	0.804	0.867	0.456
Train	e_p	551	0.957	0.940	0.980	0.593	0.933	0.915	0.953	0.399
Train	a_f	976	0.897	0.877	0.938	0.548	0.904	0.877	0.942	0.503
Train	m_s	103	0.924	0.621	0.709	0.508	0.939	0.621	0.706	0.467
Validation	all	533	0.913	0.821	0.873	0.542	0.912	0.819	0.876	0.475
Validation	e_p	140	0.963	0.920	0.973	0.659	0.955	0.912	0.975	0.478
Validation	a_f	354	0.938	0.901	0.966	0.539	0.941	0.904	0.970	0.501
Validation	m_s	39	0.839	0.641	0.681	0.429	0.840	0.641	0.681	0.444
Test	all	251	0.964	0.935	0.979	0.638	0.966	0.937	0.979	0.573
Test	e_p	73	0.986	0.959	0.990	0.633	0.986	0.959	0.990	0.458
Test	a_f	159	0.972	0.899	0.981	0.641	0.978	0.906	0.983	0.606
Test	m_s	19	0.933	0.947	0.964	0.640	0.933	0.947	0.964	0.653

tab. 1b. Comparative training, validation and held-out test metrics for the YOLOv8n-seg model. Metrics are reported globally and per class for both bounding-box detection and mask segmentation outputs.

At the global level, the model achieved a bounding box precision of 0.913, a recall of 0.821, and an mAP50 of 0.873; segmentation outputs exhibited similarly high accuracy, with mask precision and recall values of 0.912 and 0.819, respectively, and an mAP50 of 0.876. These results indicate that the network effectively identifies and delineates cartographic symbols despite their small size and the heterogeneous graphical structure of historical map sheets. Performance on the training split (Box mAP@50 = 0.876; Mask mAP@50 = 0.867) is essentially indistinguishable from that obtained on the validation set (0.873 and 0.876 respectively), and for some classes and metrics the validation scores are marginally higher. This near-perfect agreement between in-sample and out-of-sample performance indicates that the network has not memorised the training instances and that its representation of cartographic symbols generalises across map sheets.

Class-wise results further highlight the robustness of the model for the two most represented classes, *elemento_puntuale (e_p)* and *area_frammenti (a_f)*, demonstrating reliable localisation and segmentation of compact and mid-scale graphical elements.

optimistic performance estimates that would arise from computing detection metrics on the same data used to optimize the model. Training-set metrics were therefore computed *a posteriori* by applying the standard YOLOv8 evaluation procedure to the training split (84 images, 1630 annotated instances) in inference mode, using the same weights, input resolution and confidence/IoU thresholds adopted for the validation and test sets, so that the three splits are directly comparable.

Conversely, the *materiale_sporadico (m_s)* class, characterised by fewer training examples, yielded lower yet still meaningful scores, reflecting its intrinsic complexity within dense topographic backgrounds[30].

Inference speed was also favourable: preprocessing averaged 1.7 ms per image, model inference 14.7 ms, and post-processing 29.2 ms. This performance confirms that the trained model is suitable for large-scale or batch processing of historical maps in automated contexts. All validation outputs, including predicted masks and evaluation curves, were retained for subsequent comparison with ground-truth annotations and for integration into the post-training GIS pipeline.

Test Inference

After training, the best checkpoint was selected for downstream evaluation and then applied to the held-out test set to generate segmentation predictions. During this phase, all outputs are automatically exported as *.txt* annotation files in YOLO segmentation format, thereby providing the structured data required for subsequent analyses.

Quantitative evaluation on the held-out test set was performed using the standard YOLOv8 validation procedure. A confidence threshold of 0.25 was applied only during the prediction and export stage used for qualitative visualisation of detections, in order to retain plausible but uncertain predictions given the small size, irregular morphology, and variable visibility of cartographic symbols.

Test-set results indicate robust performance under the current experimental conditions, although some differences emerged between validation and test metrics across classes, particularly for the underrepresented *materiale_sporadico* category (validation mAP50 = 0.681; test mAP50 = 0.964). These differences may partly reflect the limited size and non-uniform composition of the available data and should therefore be interpreted with caution (**tab. 1b**).

Post-inference Analysis of Predictions

A visual assessment of the correspondence between predicted outputs and GT annotations was conducted on the whole subset of test images. Qualitative inspection highlights only a limited number of false negatives, typically isolated to individual instances and this observation is consistent with the high recall values obtained in quantitative evaluation (**fig. 2**).

As qualitative inspection alone cannot capture the model's internal confidence structure, the visual assessment was complemented by a post-inference quantitative

[30] Interestingly, from an archaeological perspective, one might expect *area_frammenti* to be the most challenging class to detect, given its greater variability in size, density, and internal point distribution. *Materiale_sporadico*, despite the limited number of training examples, is instead represented by a more standardised graphical convention. This is reflected in the metrics: the class achieves high precision (P = 0.924 on the training split) but comparatively low recall (R = 0.621), indicating that its graphical signature is distinctive enough to avoid false positives, while the limited number of annotated examples constrains the model's ability to retrieve all occurrences.

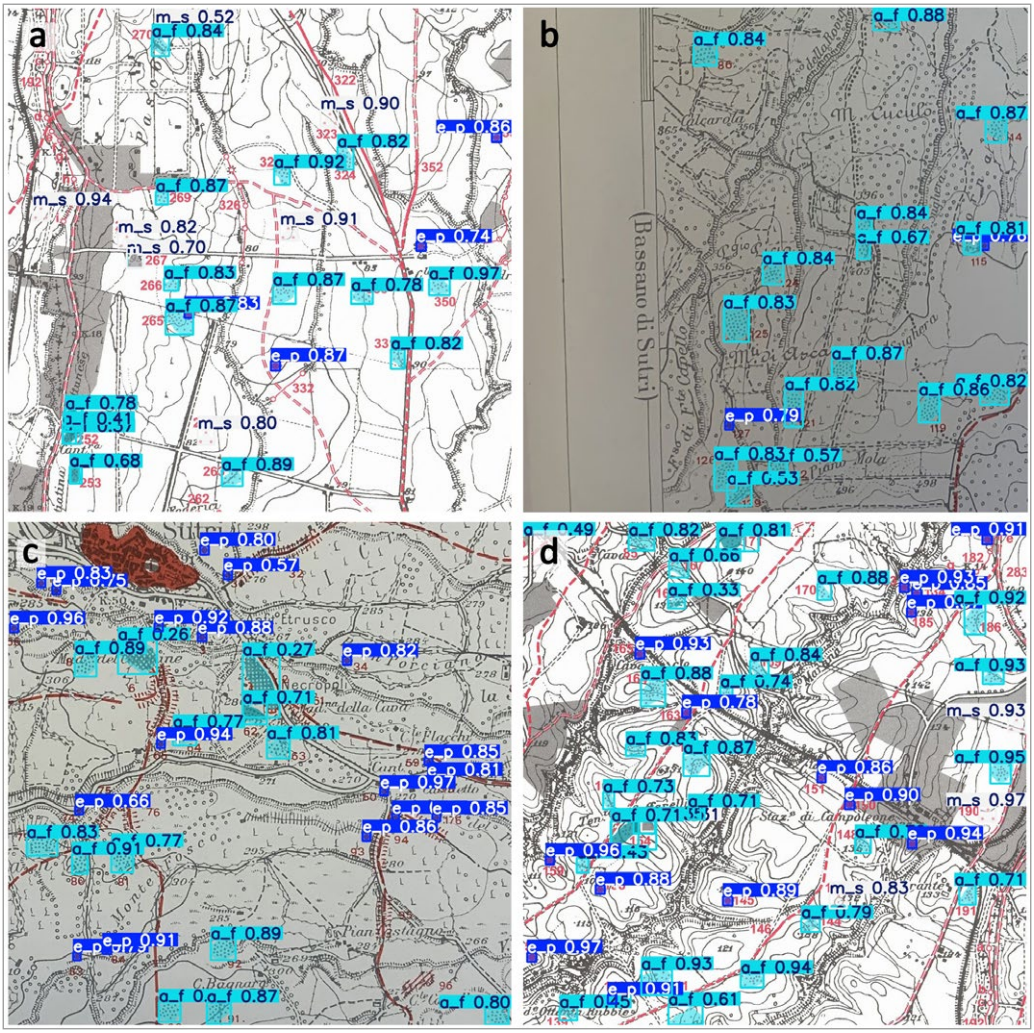

fig. 2. Representative examples from the test set showing (a, b) correct detections, (c) a false positive caused by visually similar cartographic elements, and (d) a false negative occurring in a visually dense context. The label associated with each feature reports both the class and the corresponding confidence value.

analysis, aimed at examining prediction certainty and class-level confidence patterns across the test-set.

The extracted predictions show a class distribution consistent with both the sample composition and the quantitative evaluation results, with rarer classes producing fewer detections. Confidence scores are generally high, with a median value above 0.8, indicating that most predictions are assigned a high level of confidence by the model. Lower confidence values are restricted to a small subset of detections close to the applied threshold, suggesting that uncertain cases are appropriately constrained.

The confidence score distribution is strongly skewed towards high values, with most predictions concentrated above 0.7 and a median above 0.8. Lower confidence

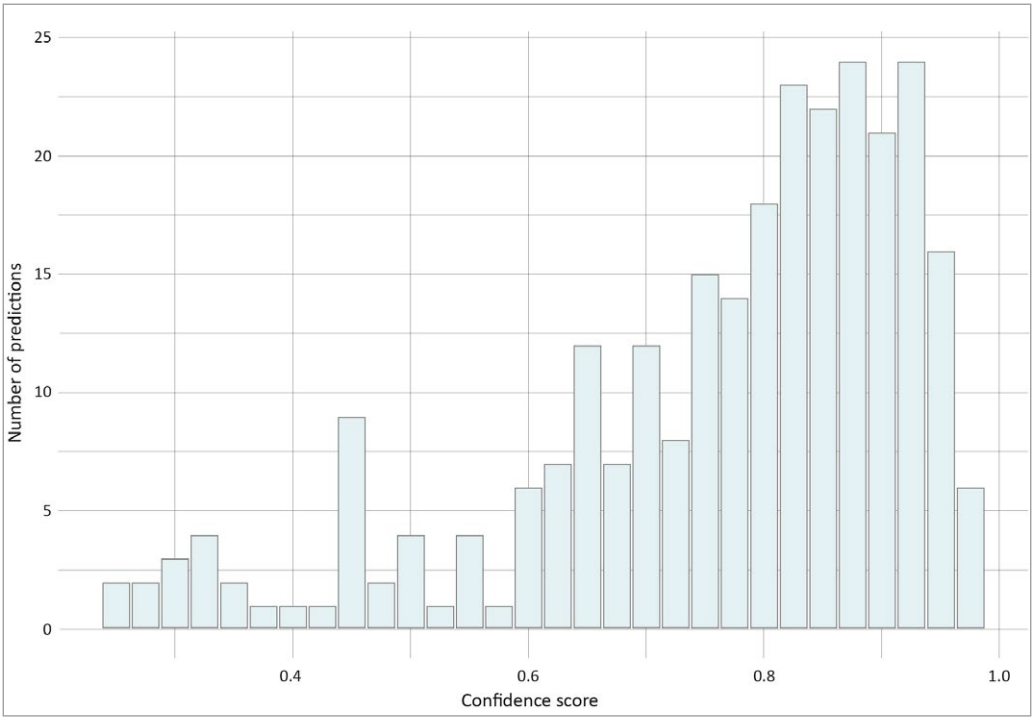

fig. 3. Distribution of confidence scores for test-set predictions. Most detections are associated with high confidence values, with only a small fraction of predictions close to the confidence threshold.

	elemento_puntuale (e_p)	**area_ frammenti (a_f)**	**materiale_ sporadico (m_s)**
number of predictions	74	168	30
mean confidence	0.848	0.746	0.656
median confidence	0.888	0.784	0.741
standard deviation of confidence	0.129	0.152	0.248
first quartile (25th percentile)	0.824	0.663	0.382
third quartile (75th percentile)	0.93	0.857	0.887

tab. 2. Class-level summary of confidence scores. The table reports the number of predictions and descriptive statistics of the confidence scores for each detected class.

values are limited to a narrow range close to the applied threshold, indicating that uncertain detections are relatively rare and that most predictions are assigned relatively high confidence scores by the model (**figs. 3-5**).

A per-class analysis of confidence scores further clarifies model behaviour. Differences emerge in both central tendency and dispersion, reflecting class-specific patterns in prediction confidence. To support the interpretation of the distribution plots, a class-level summary of confidence scores was computed (**tab. 2**): confidence distributions differ across classes, with *elemento_puntuale* showing the highest median confidence and *materiale_sporadico* the greatest variability.

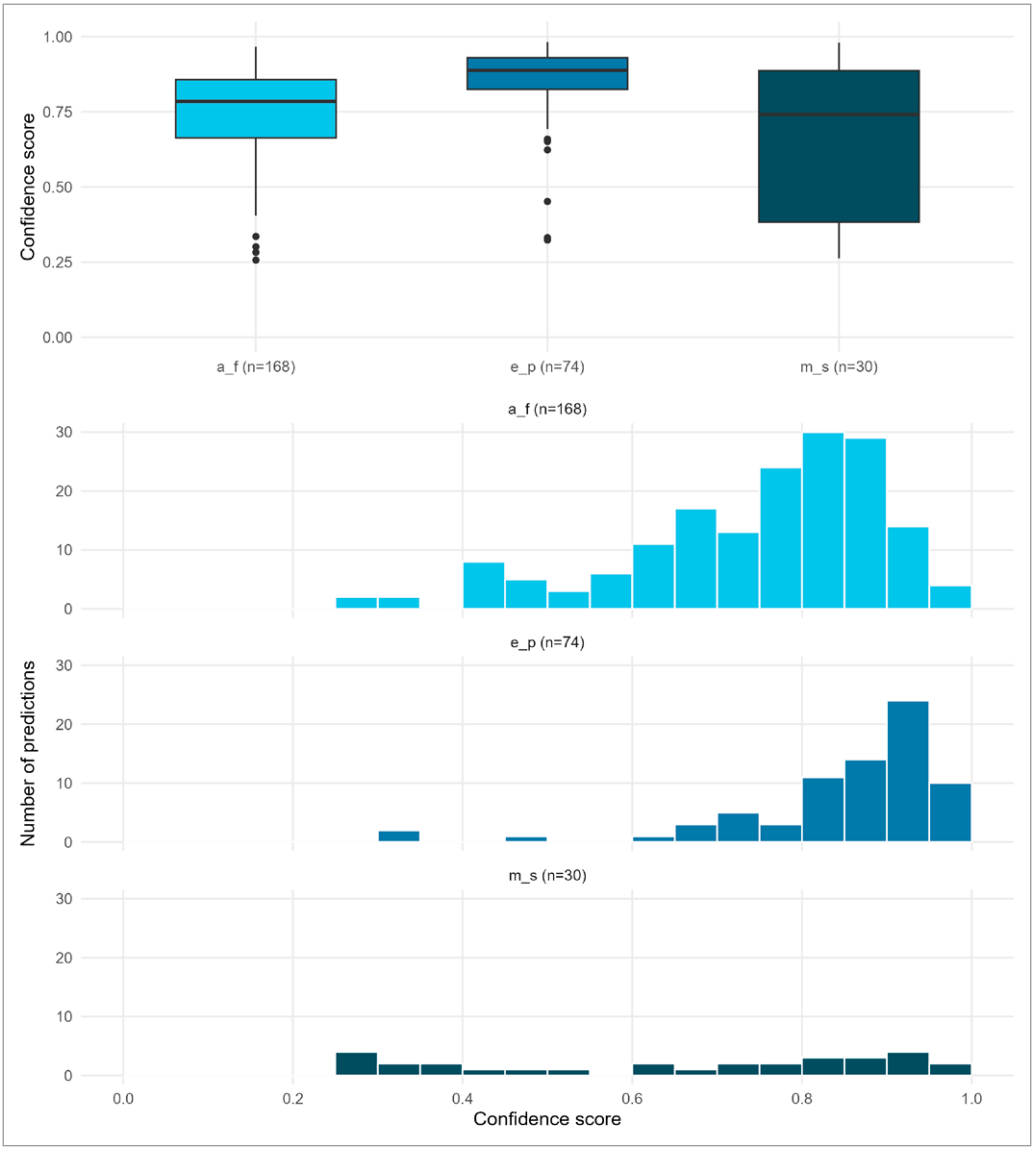

fig. 4. Boxplots show the distribution of confidence scores for each symbol class, highlighting differences in central tendency and dispersion. More frequent and visually distinctive classes are associated with higher and more stable confidence values, whereas rarer and morphologically heterogeneous symbols exhibit lower median confidence and greater variability.

5. Post-training workflow: from detection to GIS-ready data

The primary objective of the project was to extract GIS-ready features, offering substantial practical advantages over manual digitisation. Automated extraction drastically reduces processing time by generating large numbers of features in a single inference step, while ensuring methodological standardisation through the consistent

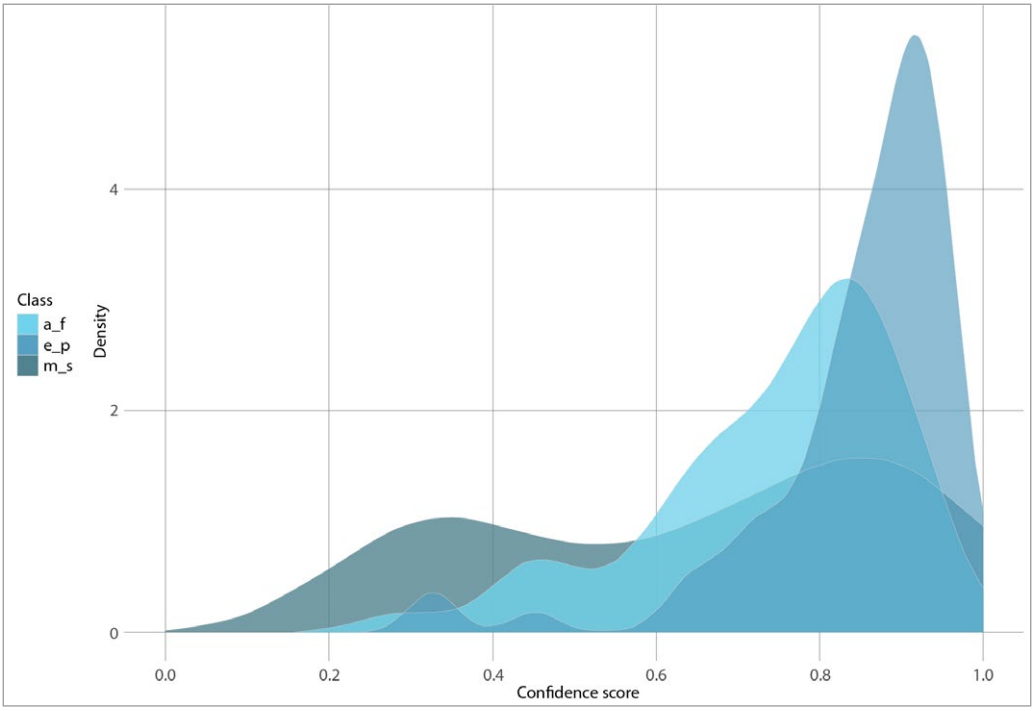

fig. 5. Distribution of confidence scores across the classes. Density curves show distributions generally skewed towards high-confidence detections. Differences in the shape and spread of the distributions reflect variations in class frequency, morphological heterogeneity, and symbol recognisability.

application of the same recognition rules across maps. This approach also enables scalability, allowing dozens or even hundreds of maps to be processed, an otherwise unfeasible task with manual workflows. Once detections are converted into spatial coordinates, results are immediately available in tabular form, ready for direct integration into GIS environments.

Domain expert validation remains essential, but the role of the archaeologist shifts from repetitive transcription to critical review, ensuring both efficiency and interpretative reliability.

In general terms, two alternative procedures can be adopted to obtain GIS-ready data, depending on whether the model is able to manage georeferenced info. Some of the research discussed earlier rely on tools capable of natively handling spatial information, resulting in geodata being produced directly during the prediction stage. In this study, however, the use of YOLO required an additional post-processing step to convert detection outputs into geospatial data[31].

Again, the workflow was executed in Google Colab. First, full polygonal masks[32], initially expressed in pixel coordinates, were extracted as derived directly from the

[31] Obviously in order to perform this task, the maps must be previously georeferenced.

[32] Unlike detection models that only provide bounding boxes, segmentation allowed us to extract full polygonal masks corresponding to the symbols of interest.

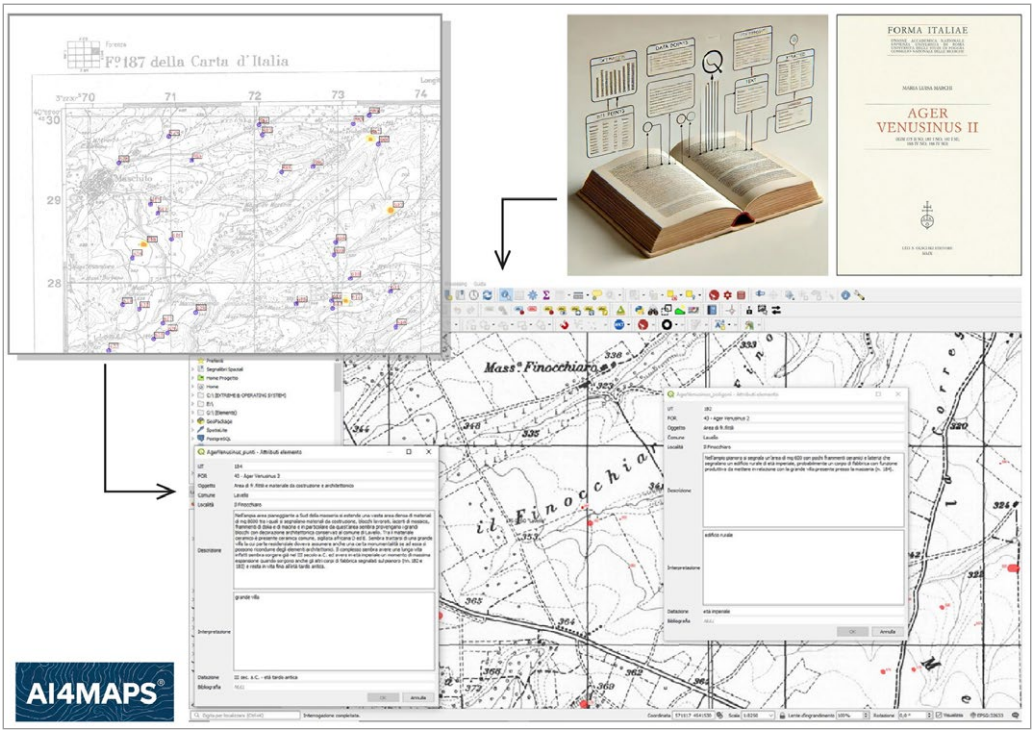

fig. 6. Data extracted using AI models are transferred into a GIS environment, enabling the semantic enrichment of spatial objects and the execution of advanced analyses.

raster image; the result was then stored in *.csv* format using *pandas*, and subsequently processed with *rasterio* and *pyproj* for affine transformations and Coordinate Reference System (CRS) management, as well as with *shapely* and *geopandas* for geometry handling and export to standard GIS format[33]. In this way, a mediated form of interoperability with GIS tools is achieved, allowing for further spatial analyses (**fig. 6**).

6. Learning Symbol Morphologies: Cropping and Classification Approach (Model B)

The high iconographic variability of archaeological symbols across historical maps, together with the frequent presence of background noise, overlapping map elements, and scanning artefacts, poses significant challenges to map-level recognition. When symbols are embedded within dense and heterogeneous cartographic contexts, these factors can hinder the reliable learning of their morphological characteristics. Training models directly on full map sheets therefore entails a high degree of complexity, as

[33] Spatial referencing was established using *world files*, which store pixel size, image extent, and positional information within a defined CRS.

archaeological maps vary widely in visual structure, making generalisation particularly difficult.

By contrast, the repertoire of archaeological symbols is comparatively limited and recurrent, despite local stylistic variations. This asymmetry motivated a shift in perspective: rather than learning objects exclusively within the cartographic context, a second, symbol-centred recognition workflow was developed to focus on them as a closed set of visual entities. The underlying assumption is that, while maps differ substantially, symbols constitute a more stable and transferable visual vocabulary. Learning their morphology in isolation thus represents a necessary step towards map-independent symbol recognition.

By removing background noise, the model is encouraged to learn discriminative features intrinsic to symbol shape. This focus also lays the methodological groundwork for future developments aimed at directly interpreting map legends, enabling the system to associate symbol forms with semantic categories (e.g., point features and fragment areas) regardless of the specific cartographic source in which they appear.

Symbol Cropping and Dataset Construction

The symbol-centred workflow builds directly upon the annotations produced during the map-level detection phase. Cropping was performed by using the bounding information associated with each annotated symbol exported from CVAT (v1.1 XML), ensuring that each crop closely enclosed the graphical element while minimising the inclusion of background noise. This operation transformed the original map-level annotations into a symbol-level image dataset suitable for classification tasks, which was split into training, validation, and test subsets[34].

The extracted crops were also organised into class-specific directories corresponding to the same semantic categories defined during annotation.

Symbol-level Data Augmentation and Classification Training

A dedicated data augmentation strategy was applied to the training subset of the symbol crops in order to increase intra-class variability. Unlike whole-map augmentation, which must preserve global cartographic coherence, symbol-level augmentation can be more aggressive while remaining morphologically meaningful, with a combination of geometric and photometric transformations designed to simulate realistic variations: ±5° rotations, translations of up to 5%, scaling between 0.95 and 1.05, horizontal flips (50% probability), brightness and contrast variations in the 0.9-1.1 range, Gaussian noise ($\sigma \approx 0.02$), and light blur (20% probability, radius 0.6-1.2).

Symbol-level classification of the augmented dataset (10698 instances) was trained using a pretrained YOLOv8 classification backbone (YOLOv8n-cls) and fine-tuned on the cropped-symbol dataset. The training configuration was selected to balance

[34] By deriving the symbol-level dataset directly from map annotations, consistency between the two recognition pipelines was preserved, ensuring that both workflows operate on the same underlying semantic definitions.

Split	Images	Top-1 accuracy
Train	10,698	0.9999
Validation	509	1.0000
Test	258	1.0000

tab. 3. Top-1 accuracy of the classification model on the training, validation and held-out test splits. Top-1 accuracy is defined as the proportion of images for which the class assigned the highest predicted probability corresponds to the ground-truth label. All values were computed with the same evaluation procedure and the same weights. A single image was misclassified across the entire dataset.

convergence stability with the computational constraints of a lightweight cloud environment. Training was run for 50 epochs with a batch size of 32 and an early-stopping criterion was enabled with a patience value of 15. Training converged rapidly, with the best validation performance reached at epoch 15 and early stopping halted the training process after 30 epochs without further improvement, thereby containing the risk of overfitting.

While the segmentation workflow addresses the spatial localisation of symbols within complex map layouts, the classification task is conceptually distinct, focusing exclusively on recognising symbol types based on their isolated visual form. Together, the two pipelines form a complementary framework. Although they are presented as independent workflows in the present study, their complementary nature establishes the basis for future integrations, such as leveraging legend symbols as reference prototypes or combining classification outputs with map-level detections to enhance robustness and interpretability.

Results

Training curves indicate stable convergence of the symbol-level classifier. Training and validation losses decrease consistently and stabilise at low values, while Top-1 validation accuracy rapidly reaches a plateau close to 1.0 and the model achieved a Top-1 test accuracy of 100%, with macro and weighted-average precision, recall, and F1-score all equal to 1.00. No evident signs of overfitting are observed, suggesting that the selected training configuration and augmentation strategy appear generally appropriate for the task (**fig. 7** and **tab. 3**)[35].

Evaluation on the held-out test set confirms the strong separability of the classes. The confusion matrix, computed on the independent test, shows all predictions correctly assigned to their respective classes, with no observed confusion between categories (**fig. 8**). The apparent discrepancy with the validation accuracy is expected, as the latter was computed on a different dataset during training, whereas the confusion matrix reflects the final evaluation on the independent test set. This result reflects the

[35] Since Ultralytics YOLO computes classification metrics on the validation set at the end of each training epoch, rather than on the training set, **fig. 7** reports the evolution of the model's validation accuracy and validation loss during training. This approach provides a more reliable assessment of the model's generalisation performance.

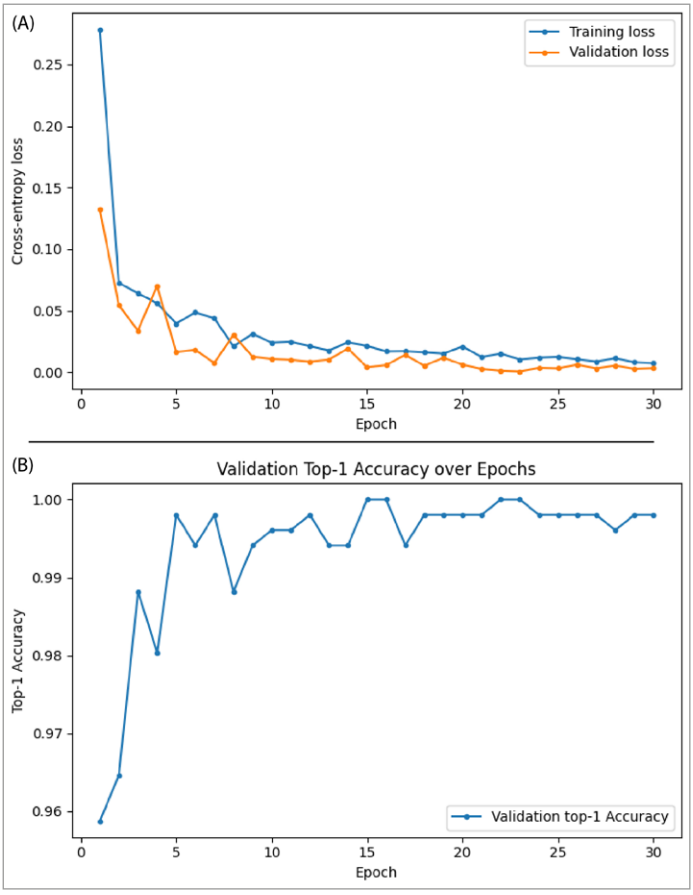

fig. 7. Training behaviour of the classification model. (a) Training and validation cross-entropy loss for symbol-level classification. (b) Top-1 accuracy on the validation set, defined as the proportion of images whose highest-scoring predicted class corresponds to the ground-truth label (truncated vertical axis, 0.95-1.00).

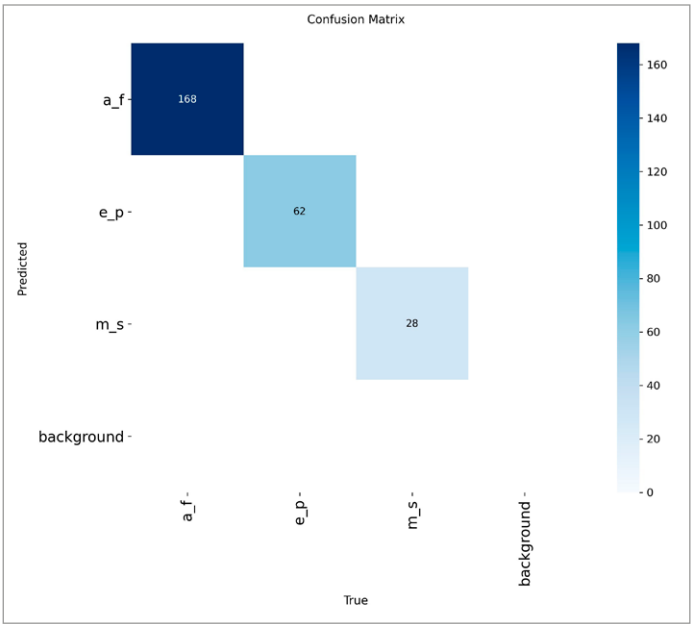

fig. 8. Confusion matrix for symbol-level classification on the test set. All symbols are correctly classified, reflecting the high separability of symbol morphologies in the controlled symbol-only setting.

controlled nature of the symbol-only setting and the morphological distinctiveness of the considered classes. Rather than indicating triviality, the absence of misclassifications demonstrates that isolating symbols from their cartographic context enables the model to learn stable and discriminative visual representations. These findings support the role of symbol-level classification as a foundation for future recognition strategies.

7. Lateral Thinking with *legendae*: towards prototype-driven recognition

Building on the idea of symbols as a closed and transferable visual vocabulary, a prototype-driven recognition strategy based on map legends is currently under active development and testing and represents a promising direction for future research aimed at achieving map-independent and legend-aware symbol recognition.

Map legends act as explicit catalogues of the symbols used in each sheet. By extracting symbols from legends, it is possible to obtain a new dataset that represent the intended visual form of each category (e.g., villa, road segment, necropolis, etc.) and can serve as reference exemplars during training or inference, reducing the need to collect exhaustive annotated samples from maps.

Within this framework, legend symbols could be transformed into latent feature representations and used as anchors against which candidate detections extracted from maps are compared. Rather than relying exclusively on memorised training examples, classification would be performed by measuring similarity between candidate embeddings and prototype representations. This approach has the potential to improve the recognition of rare or poorly represented symbols and increase interpretability by explicitly linking each classification outcome to a corresponding legend reference. Conceptually, the workflow will be articulated in four main steps: (1) symbols will be cropped from map legends and encoded into prototype embeddings; (2) symbols

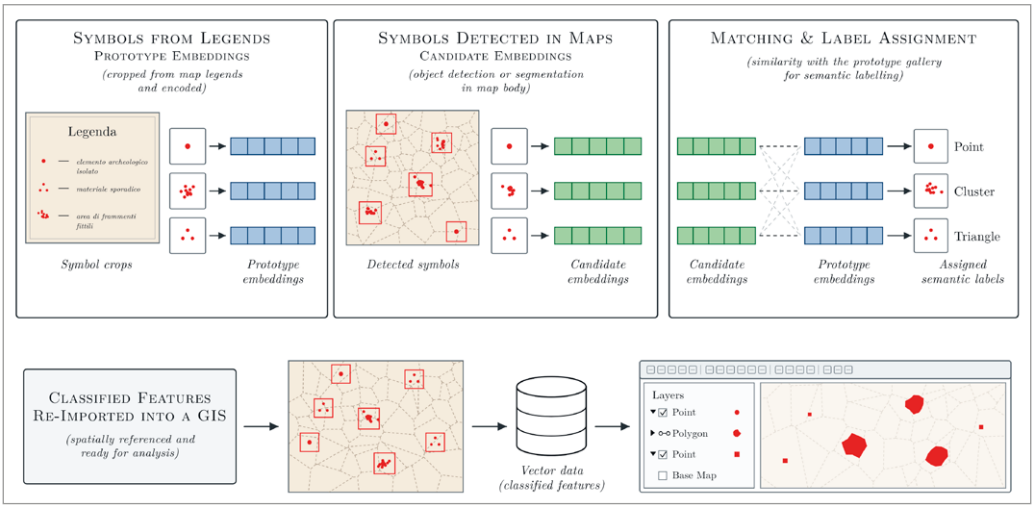

fig. 9. Proof-of-concept workflow for the automatic classification of cartographic symbols in archaeological maps using semantic embeddings.

detected within the map body through object detection or segmentation will be converted into candidate embeddings; (3) candidate embeddings will be matched against the prototype gallery to assign semantic labels based on similarity; and (4) classified features will be re-imported into a GIS environment (**fig. 9**).

8. Discussion & Ethical Implications

Limits, benefits and broader considerations

This study demonstrates that the automated recognition of archaeological symbols on historical maps can be effectively addressed through a multi-level strategy that combines map-level segmentation with symbol-level classification. Rather than relying on a single monolithic model, the proposed workflow decomposes the problem into complementary tasks, each deliberately designed to address distinct sources of complexity.

The quantitative results indicate that both segmentation and classification tasks achieve robust performance under controlled conditions. At the map level, segmentation enables the localisation and retrieval of archaeological features embedded within dense and heterogeneous cartographic contexts, maintaining high reliability despite limited training data through on-the-fly augmentation. At the symbol-level, classification results indicate that isolating symbols from their cartographic background allows the model to learn a stable and transferable visual vocabulary, supporting the idea that symbols constitute a relatively closed visual system across different authors and traditions. Together, the two approaches form also a flexible framework that can be extended, recombined, or partially reused depending on data availability and research objectives.

Beyond technical results, the study highlights a fundamental challenge inherent to archaeological cartography: the long-term evolution and variability of graphic conventions. Since the same feature may be represented through multiple symbols, training strategies must expose models to heterogeneous renderings rather than to a single fixed convention This issue encourages the learning of symbolic concepts beyond superficial appearance, promoting a mode of interpretation closer to human cartographic reading practices. Although the present experimentation is limited to three classes, the proposed solution shows clear potential for extension to a wider symbolic repertoire.

A further methodological limitation concerns class imbalance within the dataset, particularly the underrepresentation of the *materiale_sporadico* class relative to the other categories. Although this imbalance reflects the actual distribution of symbols within the available cartographic corpus, it may favour the learning of more frequent classes while reducing robustness for rarer and morphologically variable categories. This issue may also contribute to optimistic performance estimates in lightweight experimental settings with relatively small datasets. Future developments will therefore explore more targeted balancing strategies, including class-specific augmentation, synthetic data generation, and other techniques aimed at improving the representation and generalisation capacity of underrepresented classes.

From a practical perspective, the workflow significantly reduces processing time compared to manual digitisation, once the initial data preparation phase is completed and its benefits become particularly evident at larger scales.

Finally, the reliance on freely accessible tools also reflects the practical constraints under which much archaeological research is conducted. Computational resources, time, and funding are often limited, particularly outside large research infrastructures. In this context, the use of open and widely available solutions is not merely a pragmatic choice, but a methodological one. The workflow was designed as lightweight and modular, prioritising adaptability and transparency over computational intensity. Developed to operate under realistic research conditions rather than ideal laboratory settings, the approach enhances transferability and long-term sustainability, positioning AI as a supportive and context-aware instrument for archaeological practice.

Ethical Implications

The use of AI for the objectives described above raises ethical concerns that must be explicitly addressed, and that also acquire relevance within the broader debates developed in GeoAI ethics, since archaeological cartography and spatial analysis belong to the wider domain of geospatial application. In this context, recent debates in GeoAI have further stressed that spatial data present specific critical issues linked to contextuality and spatial autocorrelation, properties that may amplify geographical bias, uneven representation, and misleading cartographic interpretations (Yoo, 2026). In archaeology, these concerns are especially relevant, as maps themselves are interpretative construction rather than neutral representations[36].

One of the main issues concerns the use of incomplete, uneven, or historically situated datasets, which may reproduce distortions that, once embedded in GIS outputs, risk acquiring an unwarranted aura of objectivity. Likewise, undetected features may be implicitly interpreted as absent, reinforcing false negatives as meaningful absences within spatial analyses. Transparency and interpretability therefore remain central concerns, especially in archaeology, where methodological accountability and reproducibility are essential[37].

A further risk lies in the over-reliance on automated outputs without adequate expert human validation. Without critical assessment by archaeologists and heritage specialists, AI outputs can be misinterpreted or misapplied, leading to inappropriate reconstructions or management decisions. More importantly, they may be reused in contexts far removed from their original interpretative setting. To mitigate this risk, AI-assisted workflows must be conceived as decision-support systems rather than

[36] Increasing attention has been devoted to the non-neutral nature of geospatial technologies and GeoAI systems, as cartographic representations may reinforce asymmetries, biases, and forms of spatial exclusion while simultaneously acquiring persuasive authority through visualisation processes (Oluoch, 2024).

[37] More broadly, recent GeoAI frameworks have increasingly stressed the need for explainable AI, reproducible workflows, metadata standardisation, ethical auditing, and explicit governance mechanisms in order to mitigate the propagation of bias and support trustworthy and accountable geospatial AI systems (Pirasteh *et al.*, 2026).

autonomous interpretative tools, ensuring continuous human oversight and contextual interpretation. In this respect, increasing attention has emerged from different fields toward human-in-the-loop models, provenance tracking practices, and explicit forms of uncertainty documentation throughout the entire analytical workflow (Yoo, 2026). These perspectives, widely discussed also within GeoAI debates, highlight the need to maintain traceability between original cartographic data, AI-detected features, and subsequent interpretative steps, in order to limit the uncontrolled propagation of errors or interpretative distortions[38].

Additional ethical considerations concern data sensitivity, cartographic authority, and professional agency. The dissemination of spatial information may expose vulnerable archaeological sites to increased risks if not properly managed. At the same time, recent studies have highlighted the persuasive authority of maps and the ethical implications of cartographic visualisation itself, since AI-generated spatial outputs may encourage uncritical interpretations if uncertainty, provenance, and processing history are not adequately documented (Kang *et al.*, 2023; Kang *et al.*, 2024). In the specific case of archaeological cartography, datasets themselves may contain historical biases linked to survey intensity, preservation conditions, research traditions, or uneven territorial documentation. Consequently, AI models trained on such materials may inadvertently reproduce or reinforce pre-existing spatial asymmetries, making continuous critical assessment and expert validation indispensable (Yoo, 2026).

Excessive automation may also contribute to forms of professional de-skilling if critical interpretative roles are marginalised[39]. Responsible adoption therefore requires careful control of data access, explicit documentation of processing steps, and the preservation of expert judgement at all stages of analysis.

In this respect, the AI Act provides a relevant regulatory basis for reflecting on the responsible use of AI in cultural heritage contexts[40]. Although the experimental setup presented here does not fall within the category of high-risk AI systems, the regulation's emphasis on transparency, traceability, and documentation aligns closely with the methodological principles adopted in this study.

9. Final remarks

Future developments of the AI4MAPS project will focus on expanding the dataset, in order to progressively move towards more general frameworks adaptable to different cartographic traditions. In this perspective, achieving meaningful generalisation requires the aggregation of large datasets, which is often difficult given the limited accessibility and digitisation of historical maps. For this reason, collaborative

[38] Recent GeoAI frameworks also emphasise the importance of documentation tools such as Data Cards and Model Cards to ensure transparency, auditability, and computational reproducibility (Yoo, 2026).

[39] For a general discussion of the deskilling issue associated with AI, see Ferdman, 2025.

[40] Regulation (EU) 2024/1689 (https://artificialintelligenceact.eu/, retrieved December 10th, 2025).

initiatives aimed at data collection, annotation, and standardisation would be particularly valuable, as they could enable the integration of diverse sources and support the development of more robust and transferable models.

Since the primary objective of the project is the generation of spatial features for GIS ingestion, ongoing and future work will deepen concept-driven classification strategies. In this perspective, symbols, as instances of explicitly defined archaeological concepts, would be linked to controlled vocabularies, semantic identifiers, and ontological categories, enabling their interpretation within a broader conceptual and spatial framework.

Finally, sustainability must be acknowledged as a core consideration for future research. The energy and environmental costs associated with training and deploying AI systems require a critical assessment of when and to what extent recourse to AI is justified. In this context, the adoption of lightweight models, selective automation, and responsible use of computational resources offers a pragmatic way to ensure that AI remains a proportionate tool. AI should not be regarded as an end in itself, but as a methodological choice shaped by research questions, scale of analysis, available resources, and urgency: when applied judiciously, it can contribute to a more efficient understanding of territory and landscape dynamics, while remaining one instrument within the archaeologist's methodological toolkit.

Acknowledgements

The authors would like to thank the reviewers for their valuable comments and constructive suggestions, which significantly contributed to improving the quality and clarity of the manuscript.

Contribution

Concept and methodology: Caterina Paola Venditti, Antonio Leopardi; Data collection and management: Caterina Paola Venditti, Antonio Leopardi; Data analysis: Caterina Paola Venditti, Antonio Leopardi; Writing and revision of the original draft: Caterina Paola Venditti, Antonio Leopardi; Figures: Caterina Paola Venditti, Antonio Leopardi; Project administration: Caterina Paola Venditti, Antonio Leopardi.

Declaration of Generative AI Use

During the preparation of this work, the authors used generative AI tools solely to enhance readability and language and to support the preparation of **figs. 1** *and* **9**. *The use of these tools did not influence the scientific content, data interpretation, or conclusions of the work. The authors reviewed and edited all AI-assisted outputs as needed and take full responsibility for the content of the work.*

References

Cao, D., Yan, X., Li, J., & Wu, L. (2025). Automated Icon Extraction from Tourism Maps: A Synergistic Approach Integrating YOLOv8x and SAM. *ISPRS International Journal of Geo-Information, 14*(2), 55, 1-22. https://doi.org/10.3390/ijgi14020055

Chiang, Y.-Y., Leyk, S., & Knoblock, C.A. (2014). A survey of digital map processing techniques. *ACM Computing Surveys*, *47*(1), 1-44. https://doi.org/10.1145/2557423

Ferdman, A. (2025). AI deskilling is a structural problem. *AI & Soc*, *41*, 3001-3013. https://doi.org/10.1007/s00146-025-02686-z

Fylaktos, P., Petropoulos, G.P., & Lemesios, I. (2025). A Deep Learning Method for the Automated Mapping of Archaeological Structures from Geospatial Data: A Case Study of Delos Island. *ISPRS International Journal of Geo-Information*, *14*(6), 220, 1-15. https://doi.org/10.3390/ijgi14060220

García-Molsosa, A., Orengo, H.A., Lawrence, D., Philip, G., Hopper, K., & Petrie, C.A. (2021). Potential of deep learning segmentation for the extraction of archaeological features from historical map series. *Archaeological Prospection 2021*, *28*(2), 187-199. https://doi.org/10.1002/arp.1807

Göderle, W., Macher, C., Mauthner, K., Pimas, O., & Rampetsreiter, F. (2023). AI-driven Structure Detection and Information Extraction from Historical Cadastral Maps (Early 19th Century Franciscean Cadastre in the Province of Styria) and Current High-resolution Satellite and Aerial Imagery for Remote Sensing. arXiv. https://doi.org/10.48550/arXiv.2312.07560

Göderle, W., Pimas, O., Mauthner, K., Macher, C., & Rampetsreiter, F. (2024). Deep Learning for Historical Cadastral Maps and Satellite Imagery Analysis: Insights from Syria's Franciscean Cadastre. *Digital Humanities Quarterly, 18*(2), 1-13. https://doi.org/10.63744/asjzgcbkgkyy

Goldman, M.A., Lederer, G. W., Rosera, J.M., Graham, G.E., Mishra, A., & Yepremyan, A. (2025). Extracting data from maps: Lessons learned from the artificial intelligence for critical mineral assessment competition. *Applied Computing and Geosciences*, *27*, 100274, 1-15. https://doi.org/10.1016/j.acags.2025.100274

Gonzalez, R.C., & Woods, R.E. (1992). *Digital Image Processing*, Addison-Wesley.

Hendrycks, D., & Dietterich, T. (2019). Benchmarking Neural Network Robustness to Common Corruptions and Perturbations. arXiv. https://doi.org/10.48550/arXiv.1903.12261

Hosseini, K., Wilson, D.C.S., Beelen, K., & McDonough, K. (2022). MapReader: A Computer Vision Pipeline for the Semantic Exploration of Maps at Scale. In*Proceedings of the 6th ACM SIGSPATIAL International Workshop on Geospatial Humanities (GeoHumanities '22). Association for Computing Machinery* (pp. 8-19). https://doi.org/10.1145/3557919.3565812

Jiang, T., Li, C., Yang, M., & Wang, Z. (2022). An Improved YOLOv5s Algorithm for Object Detection with an Attention Mechanism. *Electronics, 11*(16), 2494, 1-12. https://doi.org/10.3390/electronics11162494

Jiao, C., Heitzler, M., & Hurni, L. (2022). A fast and effective deep learning approach for road extraction from historical maps by automatically generating training data with symbol reconstruction. *International Journal of Applied Earth Observation and Geoinformation, 113*, 1-12. https://doi.org/10.1016/j.jag.2022.102980

Kang, Y., Zhang, Q., & Roth, R. (2023). The Ethics of AI-generated Maps: A Study of DALL·E 2 and Implications for Cartography. arXiv. https://doi.org/10.48550/arXiv.2304.10743

Kang, Y., Gao, S., & Roth, R.E. (2024). Artificial intelligence studies in cartography: a review and synthesis of methods, applications, and ethics. *Cartography and Geographic Information Science*, *51*(4), 599-630. https://doi.org/10.1080/15230406.2023.2295943

Kim, J., Li, Z., Lin, Y., Namgung, M., Jang, L., & Chiang, Y. (2023). The mapKurator System: A Complete Pipeline for Extracting and Linking Text from Historical Maps. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '23). Association for Computing Machinery, Article 35* (pp. 1-4). https://doi.org/10.1145/3589132.3625579

Knowledge Computing Lab, Department of Computer Science & Engineering, University of Minnesota. (2025). MapKurator GitHub. Retrieved 4 December 2025. https://github.com/knowledge-computing/mapkurator-system

Kramm, T., Nyamari, N., Moseti, V., Klee, A., Vehlken, L., Anderson, D.M., Bogner, C., & Bareth, G. (2025). Deep learning-based extraction of Kenya's historical road network from topographic maps. *Sci Data, 12*, 1149, 1-14. https://doi.org/10.1038/s41597-025-05442-6

Liu, X., Gong, W., Shang, L., Li, X., & Gong, Z. (2023). Remote Sensing Image Target Detection and Recognition Based on YOLOv5. *Remote Sensing, 15*(18), 4459, 1-24. https://doi.org/10.3390/rs15184459

Mai, G., Xie, Y., Jia, X., Lao, N., Rao, J., Zhu, Q., Liu, Z., Chiang, Y.-Y., & Jiao, J. (2025). Towards the next generation of Geospatial Artificial Intelligence. *International Journal of Applied Earth Observation and Geoinformation*, *136*, 104368, 1-20. https://doi.org/10.1016/j.jag.2025.104368

Marchi, M.L. (2023). Carte archeologiche e censimenti di beni culturali. L'esperienza delle Università. *Bollettino di archeologia online XIV*, 2023/Supplemento 1, 15-35. https://doi.org/10.60978/BAO_XIV_Suppl_01_01

Martinez, T., Hammoumi, A., Ducret, G., Moreaud, M., Deschamps, R., Piegay, H., & Berger, J. F. (2023). Deep learning ancient map segmentation to assess historical landscape changes. *Journal of Maps*, *19*(1), 1-11. https://doi.org/10.1080/17445647.2023.2225071

McCarthy, A. (2025). A Scalable Machine Learning Pipeline for Building Footprint Detection in Historical Maps. arXiv. https://doi.org/10.48550/arXiv.2508.03564

Migliorati, L., & Leopardi, A. (forthcoming). *Gian Francesco Gamurrini e la Carta Archeologica d'Italia*, in Atti del Convegno *"Gian Francesco Gamurrini (1835-1923). L'Archeologo e la sua eredità a 100 anni dalla morte"* (Milano, 21-22 ottobre 2024).

O'Hara, R., Marwaha, R., Zimmermann, J., Saunders, M., & Green, S. (2024). Unleashing the power of old maps: Extracting symbology from nineteenth century maps using convolutional neural networks to quantify modern land use on historic wetlands. *Ecological Indicators*, *158*, 111363, 1-11. https://doi.org/10.1016/j.ecolind.2023.111363

Oluoch, I. (2024). Crossing Boundaries: The Ethics of AI and Geographic Information Technologies.*ISPRS International Journal of Geo-Information, 13*, 87, 1-27. https://doi.org/10.3390/ijgi13030087

Pirasteh, S., Mohamed-Ghouse, Z.S., Chauhan, L., Gartner, G., Shirmohammadi, M., & Khonsari, N. (2026). Towards Responsible GeoAI Frameworks for Ethical Governance and Global Sustainability. *Earth Systems, Resources, and Sustainability*, *1*(1), 18-31. https://doi.org/10.53941/esrs.2026.100002

Saxton, A., Dong, J., Bode, A., Jaroenchai, N., Kooper, R., Zhu, X., Kwark, D.H., Kramer, W., Kindratenko, V., & Luo, S. (2024). Accurate Feature Extraction from Historical Geologic Maps Using Open-Set Segmentation and Detection. *Geosciences*, *14*(11), 305, 1-17. https://doi.org/10.3390/geosciences14110305

Schlegel, I. (2021). Automated Extraction of Labels from Large-Scale Historical Maps. *AGILE: GIScience Series*, *2*(12), 1-14. https://doi.org/10.5194/agile-giss-2-12-2021

Schlegel, I. (2023). A Holistic Workflow for Semi-automated Object Extraction from Large-Scale Historical Maps. *KN - Journal of Cartography and Geographic Information*, *73*, 3-18. https://doi.org/10.1007/s42489-023-00131-z

Venditti, C.P. (2025). AI-Assisted Archaeology. Leveraging ChatGPT for Archaeological Maps Analysis. In Vaughan, C., Moldovan, I.M., Moldovan, S. & Fitzpatrick, N. (2025). Transdisciplinary perspectives on AI: The fourth annual conference of the European Culture and Technology Laboratory (Dublin 2025) (pp. 64-84). https://doi.org/10.21427/ntvj-t846

Xydas, Ch., Kesidis, A., Kalogeropoulos, K., & Tsatsaris, A. (2022). Buildings Extraction from Historical Topographic Maps via a Deep Convolution Neural Network. In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022) - Volume 5: VISAPP* (pp. 485-492). https://doi.org/10.5220/0010839700003124

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, *1168*(2), 1-6. https://doi.org/10.1088/1742-6596/1168/2/022022

Yoo, S. (2026). An Operational Ethical Framework for GeoAI: A PRISMA-Based Systematic Review of International Policy and Scholarly Literature. *ISPRS International Journal of Geo-Information*, *15*(1), 51, 1-53. https://doi.org/10.3390/ijgi15010051

Wood, R., Hosseini, K., Westerling, K., Smith, A., Beelen, K., Wilson, D.C.S., & McDonough, K. (2024). MapReader: Open software for the visual analysis of maps. *Journal of Open Source Software, 9*(101), 6434. Retrieved 4 December 2025. https://doi.org/10.21105/joss.06434

Zhang, H., Zhou, X., Li, H., Zhu, G., & Li, H. (2022). Machine Recognition of Map Point Symbols Based on YOLOv3 and Automatic Configuration Associated with POI. *ISPRS International Journal of Geo-Information, 11*(11), 540, 1-21. https://doi.org/10.3390/ijgi11110540