

Michał Jakubczak*
Michał Tyszkiewicz**
Michał Leloch***

Zastosowanie uczenia maszynowego w archeologii – metodyka i studium przypadku[†]

The application of machine learning in archaeology – methodology and case study[†]

Michał Jakubczak, Michał Tyszkiewicz, Michał Leloch, *Zastosowanie uczenia maszynowego w archeologii – metodyka i studium przypadku*, „Ochrona Zabytków” 2025, nr 2, s. 31–49.

Abstrakt

W Polsce zarejestrowano do tej pory setki tysięcy stanowisk archeologicznych, jednak wciąż znaczna ich liczba pozostaje nieodkryta. Dzięki dostępności numerycznego modelu terenu (NMT) z zasobów projektu Informatyczny System Osłony Kraju (ISOK) jest możliwe identyfikowanie obiektów archeologicznych charakteryzujących się specyficzną topografią. Celem rozpoczętego przez nas projektu było opracowanie metody automatycznego wykrywania obiektów o powtarzalnych cechach morfologicznych z wykorzystaniem metod uczenia maszynowego. W ramach niniejszego etapu badania skupiliśmy się wyłącznie na kurhanach.

Zastosowana metodyka opiera się na treningu modelu na danych referencyjnych, a następnie analizie NMT z obszaru testowego przy użyciu odpowiednio przygotowanych algorytmów. Efektem prac jest identyfikacja potencjalnych lokalizacji kurhanów

* Laboratorium Bio- i Archeometrii Instytutu Archeologii i Etnologii PAN

ORCID: 0000-0002-1655-7124

e-mail: m.jakubczak@iaepan.edu.pl

** Nvidia w Santa Clara (Stany Zjednoczone)

ORCID: 0009-0008-1734-5968

e-mail: michal.tyszkiewicz@gmail.com

*** Wydział Archeologii Uniwersytetu Warszawskiego

ORCID: 0000-0003-4916-7840

e-mail: leloch.michal@gmail.com

[†] Artykuł powstał na podstawie wystąpienia podczas II edycji międzynarodowej konferencji *ETHER – Eternal Heritage / Wieczne Dziedzictwo*, która odbyła się 5 i 6 listopada 2024 roku w Warszawie. Organizatorem wydarzenia był Narodowy Instytut Dziedzictwa, a współorganizatorem – Muzeum Historii Żydów Polskich POLIN.

[†] This article is based on a presentation given during the second edition of the international *ETHER – Eternal Heritage / Wieczne Dziedzictwo* conference, held on 5 and 6 November 2024 in Warsaw. The event was organized by the National Institute of Cultural Heritage and co-organized by the POLIN Museum of the History of Polish Jews.

i sklasyfikowanie wyników według stopnia prawdopodobieństwa trafnego oznaczenia – od bardzo wysokiego do niskiego. Weryfikacja części wyników potwierdziła skuteczność przyjętego podejścia. Uzyskane rezultaty wskazują, że współpraca ekspertów z dziedziny archeologii i uczenia maszynowego pozwala tworzyć efektywne narzędzia wspierające prospekcję archeologiczną oraz może znacząco ograniczyć liczbę miejsc wymagających weryfikacji terenowej.

Słowa kluczowe

archeologia, numeryczny model terenu, kurhany, uczenie maszynowe

Abstract

To date, hundreds of thousands of archaeological sites have been registered in Poland, but a significant number are yet to be discovered. Thanks to the availability of a Digital Terrain Model (DTM) from the resources of the Informatyczny System Osłony Kraju (ISOK) project, it is possible to identify archaeological objects characterized by specific topographic features. The aim of our project was to develop an automatic method of detecting objects with repetitive morphological features using machine learning. For this initial research stage, our focus was exclusively on burial mounds.

Our methodology involves training the model using reference data and then conducting a DTM analysis of the test area with appropriately prepared algorithms. The outcome of this work is the identification of potential burial mound locations, classified according to the probability of accurate identification, ranging from very high to low. Verifying some of the results confirmed the effectiveness of this approach. Our findings suggest that collaboration between archaeology and machine learning experts can lead to the development of effective tools for archaeological prospecting, significantly reducing the number of sites requiring on-site verification.

Keywords

archaeology, digital terrain model, burial mounds, machine learning

SZTUCZNA INTELIGENCJA (SI), W TYM W SZCZEGÓLNOŚCI UCZENIE MASZYNOWE, OD DŁUŻSZEGO czasu znajduje zastosowanie w nauce i przemyśle. Ostatnie lata przyniosły jednak wyraźny wzrost jej możliwości i skuteczności. Dzięki temu w literaturze archeologicznej obserwujemy rosnącą liczbę prac prezentujących potencjał metod opartych na sztucznej inteligencji, z naciskiem na wykorzystanie uczenia maszynowego w analizie danych teledetekcyjnych¹.

Nadmierny entuzjazm i brak krytycznego spojrzenia mogą niestety prowadzić do wyciągania błędnych wniosków. Wprowadzanie nowych technik badawczych wiąże się z wyzwaniem metodycznymi i metodologicznymi, które na wczesnym etapie adaptacji nowych rozwiązań bywają niedostrzegane lub marginalizowane. Dotyczy to zarówno sposobu przygotowania i jakości danych, jak i procedur oceny wyników oraz zakresu ich interpretacji.

Aby w pełni wykorzystać potencjał sztucznej inteligencji we współczesnej archeologii, potrzebna jest klarowna i spójna metodologia, wsparta wypracowanymi w ramach szerokiej dyskusji środowiskowej procedurami, zasadami i standardami. Wówczas narzędzia oparte na uczeniu maszynowym mogą się stać wartościowym wsparciem w pracy archeologa przy zachowaniu przejrzystości procesu badawczego i wiarygodności wniosków. Niniejszy artykuł wpisuje się w tę perspektywę, podejmując refleksję nad wyzwaniami, które napotkaliśmy w trakcie prac, i prezentując studium przypadku zastosowania uczenia maszynowego w analizie danych teledetekcyjnych.

¹ A. Argyrou, A. Agapiou, *A Review of Artificial Intelligence and Remote Sensing for Archaeological Research*, „Remote Sensing” 2022, vol. 14, iss. 23, doi.org/10.3390/rs14236000, dostęp: 18.05.2025; A.H. Jamil et al., *A Review on Deep Learning Application for Detection of Archaeological Structures*, „Journal of Advanced Research in Applied Sciences and Engineering Technology” 2022, vol. 26, iss. 1, s. 7–14, doi.org/10.37934/araset.26.1.714, dostęp: 18.05.2025.

Sieci neuronowe w archeologii

Sieci neuronowe to klasa modeli uczenia maszynowego, które aproksymują złożone zależności między danymi wejściowymi a wyjściowymi poprzez kompozycję wielu warstw prostych przekształceń. Parametry tych warstw (tak zwane wagi) są dostrajane na podstawie przykładów w procesie uczenia. W kontekście archeologii, w której dominują dane rastrowe, szczególne znaczenie mają konwolucyjne sieci neuronowe, projektowane do wykrywania wzorców przestrzennych.

Charakterystyczne dla badań archeologicznych jest generowanie dużej ilości danych, nawet jeśli dotyczą stosunkowo wąskiego problemu. Dane te są zazwyczaj opracowywane przez niewielkie, wyspecjalizowane zespoły badawcze. Naturalnym kierunkiem wydaje się zatem dążenie do automatyzacji procesu analiz rozbudowanych baz danych, co pozwoliłoby zmniejszyć nakład pracy konieczny do ich opracowania.

Metody oparte na sztucznej inteligencji pozwalają przede wszystkim na identyfikację i kategoryzację obiektów w złożonych zbiorach danych². Zastosowania te rozwijają się w archeologii od wielu lat i obejmują między innymi takie działania, jak:

- rozpoznawanie i klasyfikacja fragmentów ceramiki³ oraz materiałów zooarcheologicznych⁴,
- wsparcie prac konserwatorskich i rekonstrukcyjnych⁵,
- archeologia krajobrazu, w tym teledetekcja stanowisk z wykorzystaniem lotniczego skaningu laserowego (LiDAR)⁶ oraz zobrazowań lotniczych i satelitarnych⁷, a także analizy dynamiki osadnictwa⁸,
- analizy danych geofizycznych⁹.

Równolegle ukazują się syntezы przeglądowe omawiające rozwój SI w archeologii oraz związane z nim wyzwania¹⁰. W polskim kontekście odnotowano między innymi zastosowanie

² Ibidem.

³ F. Anichini et al., *The Automatic Recognition of Ceramics from Only One Photo. The ArchAIDE App*, „Journal of Archaeological Science: Reports” 2021, vol. 36, doi.org/10.1016/j.jasrep.2020.102788, dostęp: 18.05.2025; A. Anglisano et al., *Supervised Machine Learning Algorithms to Predict Provenance of Archaeological Pottery Fragments*, „Sustainability” 2022, vol. 14, doi.org/10.3390/su141811214, dostęp: 18.05.2025.

⁴ N. Fadzly et al., *A Deep Learning Framework for Bone Fragment Classification in Owl Pellets Using YOLOv12*, „Scientific Reports” 2025, vol. 15, doi.org/10.1038/s41598-025-15906-9, dostęp: 18.05.2025.

⁵ C. Toler-Franklin et al., *Multi-Feature Matching of Fresco Fragments*, „ACM Transactions on Graphics (TOG)” 2010, vol. 29, iss. 6, doi.org/10.1145/1882261.1866207, dostęp: 18.05.2025; E. Grilli, F. Remondino, *Machine Learning Generalisation across Different 3D Architectural Heritage*, „ISPRS International Journal of Geo-Information” 2020, vol. 9, iss. 6, doi.org/10.3390/ijgi9060379, dostęp: 18.05.2025.

⁶ D. Stott, S.M. Kristiansen, S.M. Sindbæk, *Searching for Viking Age Fortresses with Automatic Landscape Classification and Feature Detection*, „Remote Sensing” 2019, vol. 11, iss. 16, doi.org/10.3390/rs11161881, dostęp: 18.05.2025; A. Guyot, L. Hubert-Moy, T. Lorho, *Detecting Neolithic Burial Mounds from LiDAR-Derived Elevation Data Using a Multi-Scale Approach and Machine Learning Techniques*, „Remote Sensing” 2018, vol. 10, iss. 2, doi.org/10.3390/rs10020225, dostęp: 18.05.2025; N. Abate et al., *An Open-Source Machine Learning-Based Methodological Approach for Processing High-Resolution UAS LiDAR Data in Archaeological Contexts. A Case Study from Epirus, Greece*, „Journal of Archaeological Method and Theory” 2025, vol. 32, doi.org/10.1007/s10816-025-09706-8, dostęp: 18.05.2025.

⁷ H.A. Orenge et al., *Automated Detection of Archaeological Mounds Using Machine-Learning Classification of Multisensor and Multitemporal Satellite Data*, „PNAS” 2020, vol. 117, iss. 31, doi.org/10.1073/pnas.2005583117, dostęp: 18.05.2025.

⁸ J.J. Miera et al., *Large-Scale Investigations of Neolithic Settlement Dynamics in Central Germany Based on Machine Learning Analysis. A Case Study from the Weiße Elster River Catchment*, „PLOS One” 2022, vol. 17, iss. 4, doi.org/10.1371/journal.pone.0265835, dostęp: 18.05.2025.

⁹ A. Zakirov et al., *Artificial Intelligence in Archaeological Geophysics. Testing ChatGPT-4o on Magnetometer Data from Sapallitepa (Uzbekistan)*, „Archaeological Prospection” 2025, doi.org/10.1002/arp.70006, dostęp: 18.05.2025.

¹⁰ G. Gattiglia, *Managing Artificial Intelligence in Archaeology. An Overview*, „Journal of Cultural Heritage” 2025, vol. 71, doi.org/10.1016/j.culher.2024.11.020, dostęp: 18.05.2025.

architektury U-Net do segmentacji semantycznej form antropogenicznych w danych LiDAR z Puszczy Białowieskiej¹¹.

Dane wysokościowe są jednym z podstawowych źródeł informacji w trakcie badań teledetekcyjnych i mają wartość poznawczą o dużym potencjale odkryć¹². Praca z takimi danymi wiąże się jednak z wyzwaniem, między innymi niską jakością pozyskanych danych, a także znacznym nakładem czasu i pracy koniecznym do teledetekcji i analizy stanowisk archeologicznych w dużych zbiorach danych przestrzennych. Coraz większe obszary są skanowane z użyciem technologii LiDAR i rejestrowane na wysokorozdzielczych obrazowaniach satelitarnych, co wraz z napływem danych lepszej jakości pozyskiwanych w różnych porach roku w kolejnych latach wymusza ponawianie badań na podstawie nowych informacji. W związku z tym nie jest zaskoczeniem, że idea automatyzacji tego procesu od lat stanowi istotny przedmiot zainteresowania badaczy¹³.

Postępujący w niezwykłym tempie rozwój uczenia maszynowego daje możliwość budowania skutecznych platform automatyzujących teledetekcję. Użyteczność tej metody została już niejednokrotnie podkreślona w literaturze naukowej¹⁴.

Cel i zasięg badań

Podczas pisania niniejszego artykułu postawiliśmy sobie następujące cele:

- zaprezentowanie metody automatycznej detekcji obiektów o morfologii kurhanów na podstawie numerycznego modelu terenu (NMT) pozyskanego z użyciem technologii lotniczego skaningu laserowego,
- ocena skuteczności modelu na niezależnym obszarze testowym,
- wskazanie kluczowych problemów metodycznych, jakie napotkaliśmy w trakcie prac.

Prace realizowaliśmy w ramach projektu, którego celem było zbudowanie zautomatyzowanego systemu rozpoznawania w rzeźbie terenu pojedynczych kurhanów i ich skupisk (cmentarzysk kurhanowych), a następnie przygotowanie ogólnokrajowej warstwy poligonowej wskazań przeznaczonych do dalszej weryfikacji i analiz.

¹¹ P.Z. Banasiak et al., *Semantic Segmentation (U-Net) of Archaeological Features in Airborne Laser Scanning – Example of the Białowieża Forest*, „Remote Sensing” 2022, vol. 14, iss. 4, doi.org/10.3390/rs14040995, dostęp: 18.05.2025.

¹² D. Evans, *Airborne Laser Scanning as a Method for Exploring Long-Term Socio-Ecological Dynamics in Cambodia*, „Journal of Archaeological Science” 2016, vol. 74, doi.org/10.1016/j.jas.2016.05.009, dostęp: 18.05.2025; R.D. Hansen et al., *LiDAR Analyses in the Contiguous Mirador-Calakmul Karst Basin, Guatemala. An Introduction to New Perspectives on Regional Early Maya Socioeconomic and Political Organization*, „Ancient Mesoamerica” 2023, vol. 34, iss. 3, doi.org/10.1017/S0956536122000244, dostęp: 18.05.2025.

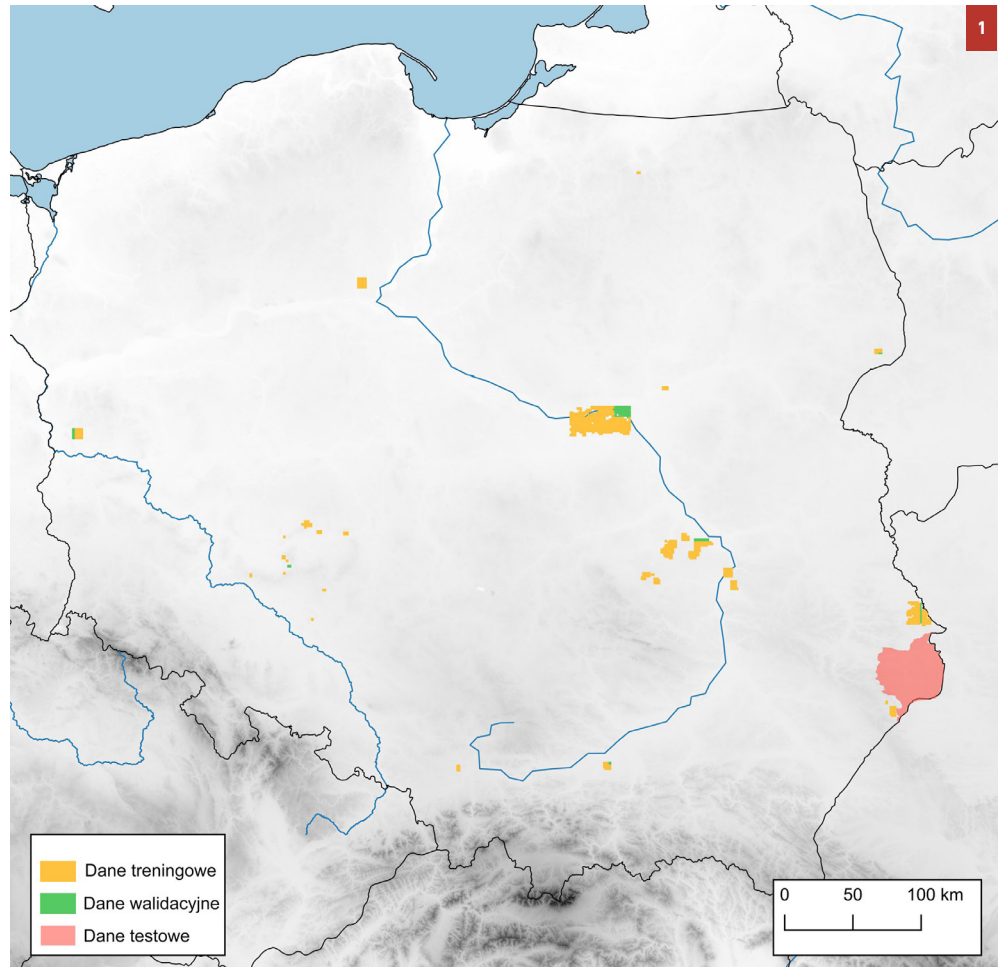
¹³ T. Freeland et al., *Automated Feature Extraction for Prospection and Analysis of Monumental Earthworks from Aerial LiDAR in the Kingdom of Tonga*, „Journal of Archaeological Science” 2016, vol. 69, doi.org/10.1016/j.jas.2016.04.011, dostęp: 18.05.2025; D.S. Davis, M.C. Sanger, C.P. Lipo, *Automated Mound Detection Using Lidar and Object-Based Image Analysis in Beaufort County, South Carolina*, „Southeastern Archaeology” 2019, vol. 38, iss. 1, doi.org/10.1080/0734578X.2018.1482186, dostęp: 18.05.2025; A. Guyot, L. Hubert-Moy, T. Lorho, *Detecting Neolithic Burial Mounds...*, op. cit.; I. Berganzo-Besga et al., *Hybrid MSRM-Based Deep Learning and Multitemporal Sentinel 2-Based Machine Learning Algorithm Detects Near 10k Archaeological Tumuli in North-Western Iberia*, „Remote Sensing” 2021, vol. 13, iss. 20, doi.org/10.3390/rs13204181, dostęp: 18.05.2025.

¹⁴ I. Berganzo-Besga et al., *Hybrid MSRM-Based Deep Learning...*, op. cit.; S. Arnoldussen et al., *Field Systems and Later Prehistoric Land Use. New Insights Into Land Use Detectability and Palaeodemography in the Netherlands through LiDAR, Automatic Detection and Traditional Field Data*, „Archaeological Prospection” 2023, vol. 30, iss. 3, doi.org/10.1002/arp.1891, dostęp: 18.05.2025; A. Karamitrou et al., *Towards the Use of Artificial Intelligence Deep Learning Networks for Detection of Archaeological Sites*, „Surface Topography. Metrology and Properties” 2022, vol. 10, iss. 4, doi.org/10.1088/2051-672X/ac9492, dostęp: 18.05.2025; S. Quintus, D.S. Davis, E.E. Cochrane, *Evaluating Mask R-CNN Models to Extract Terracing Across Oceanic High Islands. A Case Study from Sāmoa*, „Archaeological Prospection” 2023, vol. 30, iss. 4, doi.org/10.1002/arp.1909, dostęp: 18.05.2025.

1

Mapa Polski z zaznaczeniem obszarów, z których pochodziły wykorzystane w projekcie dane

Map of Poland showing the locations where the data used in the project was collected



Zakres przestrzenny zawężono do obszarów zalesionych, ponieważ w polskich warunkach większość kurhanów znajduje się pod drzewostanem, gdzie prospekcja powierzchniowa jest utrudniona, a skanowanie laserowe ma istotną przewagę. Z analiz wyłączono Karpaty i Sudety, ponieważ na tych obszarach nie udało się zgromadzić wystarczającej liczby przykładów kurhanów do uczenia sieci. Zasięg przetwarzania objął większość lasów w Polsce, natomiast zbiór treningowo-walidacyjny liczył 1237 kafli (około 0,34% powierzchni Polski), dobranych tak, aby odzwierciedlały zróżnicowanie środowisk leśnych. Niezależny obszar testowy, niewykorzystany w trenowaniu, obejmował Kotlinę Hrubieszowską, Grzędę Sokalską i Równinę Bełską (około 1850 km²) (ryc. 1). W niniejszym tekście raportujemy wyniki i przedstawiamy interpretację wyłącznie dla tego obszaru testowego, przy czym opisana metodyka dotyczy całości prac. Produktem końcowym jest ogólnokrajowa warstwa poligonowa będąca rejestrem wskazań potencjalnych kurhanów (pojedynczych i cmentarzysk), skategoryzowana według poziomu ufności modelu. Nie stanowi ona katalogu stanowisk, lecz jest narzędziem do wyboru punktów do weryfikacji terenowej oraz do dalszych analiz konserwatorskich i krajobrazowych.

Dane i ich przygotowanie

Wektorowe dane topograficzne

Źródłem wykorzystanych danych topograficznych była Baza Danych Obiektów Topograficznych (BDOT10k; tinyurl.com/yck9jcr7). Dane te są udostępniane bezpłatnie. Zawartość i szczegółowość zestawu danych w przybliżeniu odpowiada tradycyjnej mapie topograficznej w skali 1 : 10 000. Baza danych liczy obecnie około 60 milionów obiektów wektorowych. Dane wektorowe dotyczące sieci wodnej i komunikacyjnej, pokrycia terenu i budynków zostały połączone w jeden plik wektorowy.

Numeryczny model terenu

W 2010 roku w Polsce zainicjowano projekt Informatyczny System Osłony Kraju (ISOK). Jednym z jego kluczowych elementów było lotnicze skanowanie laserowe, początkowo dolin głównych rzek, a następnie obszaru całego kraju. Za źródło danych posłużyły chmury punktów pozyskane techniką lotniczego skaningu laserowego. Minimalne parametry skanowania zostały określone zgodnie z dwoma standardami. W standardzie I gęstość punktów wynosiła 4 lub 6 pkt/m². Maksymalny kąt skanowania wynosił 25° dla terenów zalesionych, przy minimalnym pokryciu 20% między pasami przelotów. Maksymalna średnica plamki lasera wynosiła 50 cm. Z kolei w standardzie II nominalna liczba punktów wynosiła 12 pkt/m². Uzyskano je podczas dwóch skanów po 6 pkt/m² wykonanych pod kątem prostym do siebie. Pozostałe parametry były analogiczne do tych opisanych dla standardu I¹⁵. Standard II zazwyczaj stosowano w obszarach miejskich, dlatego jego użyteczność w archeologii jest zdecydowanie niższa.

Od 2020 roku znaczna część danych z państwowego zasobu geodezyjnego i kartograficznego jest dostępna bezpłatnie (tinyurl.com/yc6vcx4f). Dane można pobrać za pośrednictwem standardu Web Coverage Service (WCS), który umożliwia przesyłanie do serwera zapytań dotyczących określonego obszaru. Informacja zwrotna jest dostarczana w postaci pliku w wybranym formacie i układzie odniesienia (CRS) oraz w wybranej rozdzielczości. W naszym przypadku był to plik rastrowy w formacie GeoTIFF, w Państwowym Układzie Współrzędnych Geodezyjnych 1992 (EPSG:2180) i w rozdzielczości 1 m. Otrzymaliśmy bazę danych wysokościowych dla wszystkich obszarów zalesionych w Polsce, składającą się ze 193 293 plików o łącznym rozmiarze około 750 GB.

Zastosowane metody

GIS

Podstawowym oprogramowaniem do przeglądania i analizy danych przestrzennych był QGIS. Jest to oprogramowanie typu open source, pozwalające na szybką wizualizację wyników, tworzenie i modyfikację warstw wektorowych, a także łatwe dostosowanie do konkretnych wymagań poprzez użycie wtyczek. Wizualizacje NMT wygenerowano przy użyciu oprogramowania Relief Visualization Toolbox (RVT)¹⁶.

Python

Do pracy z siecią neuronową wykorzystaliśmy język programowania Python. Umożliwia on szybkie i stosunkowo proste tworzenie aplikacji, a jednocześnie zapewnia łatwą współpracę ze specjalistycznym oprogramowaniem napisanym w językach zorientowanych na wydajność obliczeniową. Python jest obecnie najczęściej stosowanym językiem w dziedzinie uczenia maszynowego, a rozbudowany ekosystem bibliotek geoprzestrzennych pozwala sprawnie radzić sobie z danymi wysokościowymi. W projekcie został wykorzystany między innymi do automatyzacji procesu pobierania NMT, implementacji funkcji związanych z trenowaniem sieci neuronowych oraz do czyszczenia i scalania wyników.

Uczenie maszynowe

Do głębokiego uczenia zdecydowaliśmy się wykorzystać metodę segmentacji obrazów, która polega na jednoczesnej klasyfikacji wszystkich pikseli wprowadzonego rastra do zdefiniowanych wcześniej kategorii. W naszym przypadku były to kurhan i tło. Po zakończeniu klasyfikacji danego

¹⁵ Z. Kurczyński, E. Stojek, U. Cisło-Lesicka, *Zadania GUGiK realizowane w ramach projektu ISOK* [w:] *Podręcznik dla uczestników szkoleń z wykorzystania produktów LiDAR*, red. P. Wężyk, Warszawa 2014, s. 22–58.

¹⁶ K. Zakšek, K. Oštir, Ž. Kokalj, *Sky-View Factor as a Relief Visualization Technique*, „Remote Sensing” 2011, vol. 3, iss. 2, doi.org/10.3390/rs3020398, dostęp: 18.05.2025; Ž. Kokalj, M. Somrak, *Why Not a Single Image? Combining Visualizations to Facilitate Fieldwork and On-Screen Mapping*, „Remote Sensing” 2019, vol. 11, iss. 7, doi.org/10.3390/rs11070747, dostęp: 18.05.2025.

kafla sąsiadujące piksele tej samej klasy były łączone, a następnie zapisywane jako plik wektorowy. W ten sposób powstał obrys wszystkich znalezionych kopców. Metodę segmentacji obrazów wybraliśmy ze względu na szybkość jej działania i prostotę implementacji.

Podczas pracy wykorzystaliśmy konwolucyjną sieć neuronową (CNN) o architekturze U-Net¹⁷. Sieci CNN są zbudowane w taki sposób, że klasyfikacja danego piksela przez sieć zależy od zawartości ustalonego okna sąsiednich pikseli (tak zwanego pola recepcyjnego sieci), przy czym istotna jest wyłącznie zawartość tego okna, a nie jego pozycja w obrębie całego obrazu. Odpowiada to założeniu, że podział badanego obszaru na pojedyncze kafle rastrowe jest arbitralny, a uwzględnianie położenia obiektów w ramach kafla może prowadzić jedynie do powstania przypadkowych korelacji.

By zredukować potencjalne uprzedzenia kierunkowe, obecne w danych treningowych, podczas trenowania modelu stosowano losowe rotacje obrazów oraz odpowiadających im etykiet. Cztery parametry konstrukcyjne, obejmujące między innymi dobór rozmiaru pola recepcyjnego oraz liczby parametrów sieci – co odzwierciedla przybliżoną złożoność reguł wyrażanych przez model – określały zakres wpływu programisty na działanie algorytmu klasyfikacji. Reguły podejmowania decyzji przez sieć w odniesieniu do przynależności pikseli do danej klasy były zatem ograniczone do:

- niezależności od położenia w obrazie,
- zależności wyłącznie od wartości sąsiednich pikseli w obrębie pola recepcyjnego,
- niezmienniczości względem obrotów tej lokalnej struktury.

Wszystkie pozostałe cechy, takie jak wielkość, wysokość, rozgałęzienie, symetria czy współwystępowanie z innymi formami terenowymi, były w całości wyprowadzane przez model na podstawie danych treningowych.

Model był uczony nie na surowych danych wysokościowych, ale na NMT po filtracji górnoprzepustowej. Zastosowany filtr odejmuje średnią wysokość w promieniu 33 m, co pozwala zachować jedynie lokalne zróżnicowanie rzeźby terenu. Technika ta, powszechnie stosowana w archeologicznych analizach danych LiDAR (na przykład w wizualizacjach LRM), zapobiega wykorzystywaniu przez sieć neuronową dużych cech topograficznych oraz wysokości bezwzględnej jako czynników decyzyjnych.

Sieć neuronową zaimplementowaliśmy z wykorzystaniem biblioteki PyTorch¹⁸. Do obsługi danych geoprzestrzennych na etapie trenowania i wnioskowania zastosowaliśmy bibliotekę TorchGeo¹⁹. Implementacja architektury U-Net została zaś oparta na bibliotece Segmentation Models PyTorch²⁰.

Przetwarzanie danych i metodyka prac

Schemat przedstawiony na ryc. 2 ilustruje kolejne etapy prac – od momentu sformułowania pytań badawczych po publikację wyników. Proces obejmuje pięć głównych etapów. Są to:

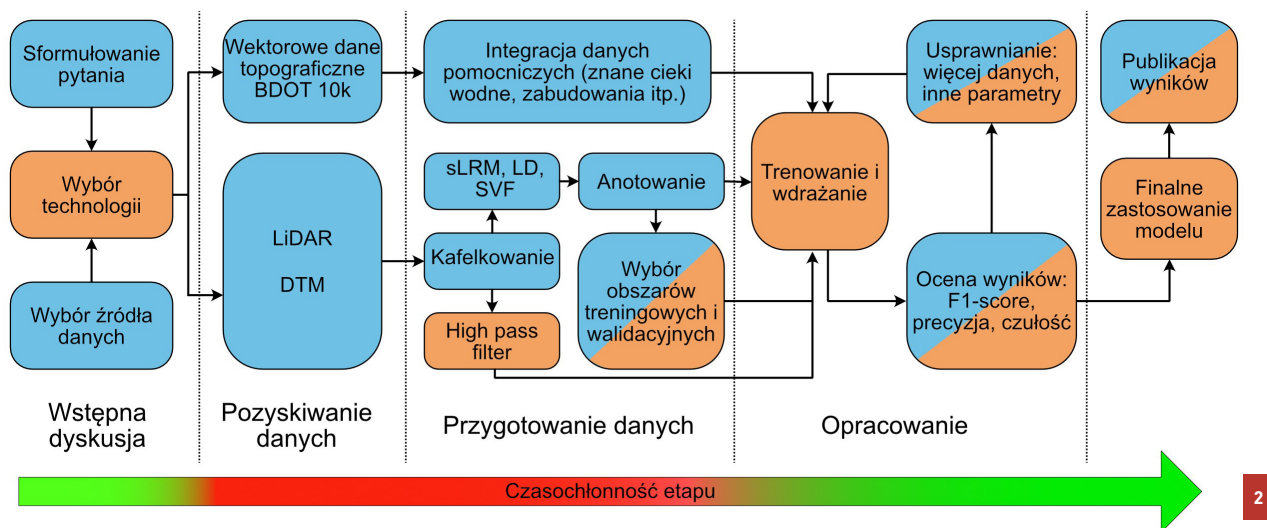
1. Wstępna dyskusja obejmująca wspólną analizę problemu badawczego przez członków zespołu, prowadzącą do sformułowania pytań badawczych oraz wyboru odpowiednich źródeł danych i technologii.

¹⁷ O. Ronneberger, P. Fischer, T. Brox, *U-Net – Convolutional Networks for Biomedical Image Segmentation* [w:] *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, „LNCS”, vol. 9351, ed. N. Navab et al., Cham 2015, doi.org/10.1007/978-3-319-24574-4_28, dostęp: 18.05.2025.

¹⁸ A. Paszke et al., *PyTorch. An Imperative Style, High-Performance Deep Learning Library* [w:] *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, ed. H. Wallach, [s.l.a.], s. 8024–8035.

¹⁹ A.J. Stewart et al., *TorchGeo. Deep Learning with Geospatial Data* [w:] *SIGSPATIAL '22. Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, New York 2022, doi.org/10.1145/3557915.3560953, dostęp: 18.05.2025.

²⁰ P. Iakubovskii, *Segmentation Models*, 2019, github.com/qubvel/segmentation_models, dostęp: 18.05.2025.



2 Schemat procesu przygotowania i wykorzystania modelu z zastosowaniem uczenia maszynowego na potrzeby rozpoznania archeologicznego. Kolorem niebieskim oznaczono zadania, za których wykonanie odpowiada archeolog, pomarańczowym – zadania specjalisty z zakresu SI. Etapy oznaczone obiema barwami wymagają współpracy specjalistów z obu dziedzin. Pod schematem oznaczono, jak zmienia się czasochłonność przedstawionych procesów na przestrzeni realizacji projektu

Diagram showing the process of preparing and using a machine learning model for archaeological reconnaissance. Tasks performed by archaeologists are marked in blue and those performed by AI specialists in orange. Stages marked with both colours require cooperation between specialists from both fields. The changes in the time required for the processes over the course of the project are indicated below the diagram

2. Pozyskanie danych, czyli identyfikacja i zebranie wymaganych zasobów (z ogólnodostępnych baz lub źródeł komercyjnych) oraz ich zapis w odpowiednich formatach i uporządkowanie w jednolitym systemie odniesienia.
3. Przygotowanie danych obejmujące czynności techniczne, takie jak kafelkowanie, filtracja górnoprzepustowa i generowanie wizualizacji, a także etap interpretacyjny – manualną analizę danych i przygotowanie anotacji, czyli oznaczeń stanowiących podstawę do trenowania modelu.
4. Opracowanie polegające na trenowaniu modelu, walidacji wyników oraz dostrajaniu parametrów, po to by osiągnąć satysfakcjonującą skuteczność.
5. Zastosowanie modelu i publikacja wyników, obejmujące integrację uzyskanych danych, przygotowanie końcowych opracowań oraz ich udostępnienie.

W dalszej części artykułu szczegółowo omawiamy przebieg kolejnych etapów, od przygotowania danych wysokościowych po proces trenowania i walidacji modelu.

Punktem wyjścia było opracowanie numerycznego modelu terenu, który następnie podzielono na kwadraty o boku 1024 m. Przygotowany został indeks w postaci pliku wektorowego, a każdy kafel otrzymał indywidualny numer identyfikacyjny. Dla danych numerycznych modeli terenu powstały wizualizacje SVF²¹, Local Dominance²² oraz SLRM, czyli uproszczona wersja wizualizacji LRM²³. Należy podkreślić, że wizualizacje te były wykorzystywane nie do trenowania modelu, ale jedynie do anotacji danych.

²¹ Ž. Kokalj, K. Zakšek, K. Oštir, *Application of Sky-View Factor for the Visualisation of Historic Landscape Features in Lidar-Derived Relief Models*, „Antiquity” 2011, vol. 85, s. 263–273.

²² Ž. Kokalj, R. Hesse, *Airborne Laser Scanning Raster Data Visualization. A Guide to Good Practice*, Ljubljana 2017, doi.org/10.3986/9789612549848, dostęp: 18.05.2025.

²³ R. Hesse, *LiDAR-Derived Local Relief Models. A New Tool for Archaeological Prospection*, „Archaeological Prospection” 2010, vol. 17, iss. 2, doi.org/10.1002/arp.374, dostęp: 18.05.2025.

Dane do anotacji zostały przygotowane w całości przez nas. Objęły one 1158 przykładów kopców interpretowanych jako kurhany. Obszary wybrane do trenowania, testowania i walidacji modelu (ryc. 1) przeanalizowało niezależnie dwóch archeologów mających doświadczenie w interpretacji danych LiDAR. Każdy kopiec został oznaczony jako punkt. Na punkty nałożono sześciometrowe bufory. Do fazy uczenia i ewaluacji wybrano łącznie 1237 kafli rozmieszczonych w całym kraju. Obszary te zostały dobrane arbitralnie, na podstawie doświadczenia członków zespołu projektowego w zakresie archeologii krajobrazu. Celem doboru było uwzględnienie terenów o zróżnicowanym charakterze krajobrazowym. Wyznaczone obszary treningowe obejmują łącznie około 0,34% powierzchni kraju. Zdecydowana większość tego obszaru może być zaklasyfikowana jako tło, gdyż stanowi ją względnie jednorodna przestrzeń pozbawiona charakterystycznych form terenowych.

By przeciwdziałać problemowi niezrównoważenia klas – wynikającemu z przewagi pikseli tła nad kurhanami – zastosowano standardowe metody, takie jak zwiększenie wag przypisanych pikselom należącym do obiektów archeologicznych oraz wykorzystanie funkcji Focal Loss²⁴.

Różne formy terenowe mogą wykazywać znaczne podobieństwo do obiektów archeologicznych, co stanowi potencjalne źródło błędów w klasyfikacji. Z tego względu na podstawie danych topograficznych opracowano warstwę maski tła. Została ona wykorzystana zarówno na etapie przetwarzania wyników – w celu odrzucenia błędnych klasyfikacji – jak i podczas uczenia sieci, aby dodatkowo zwiększyć wagę klasyfikacyjną tych obszarów.

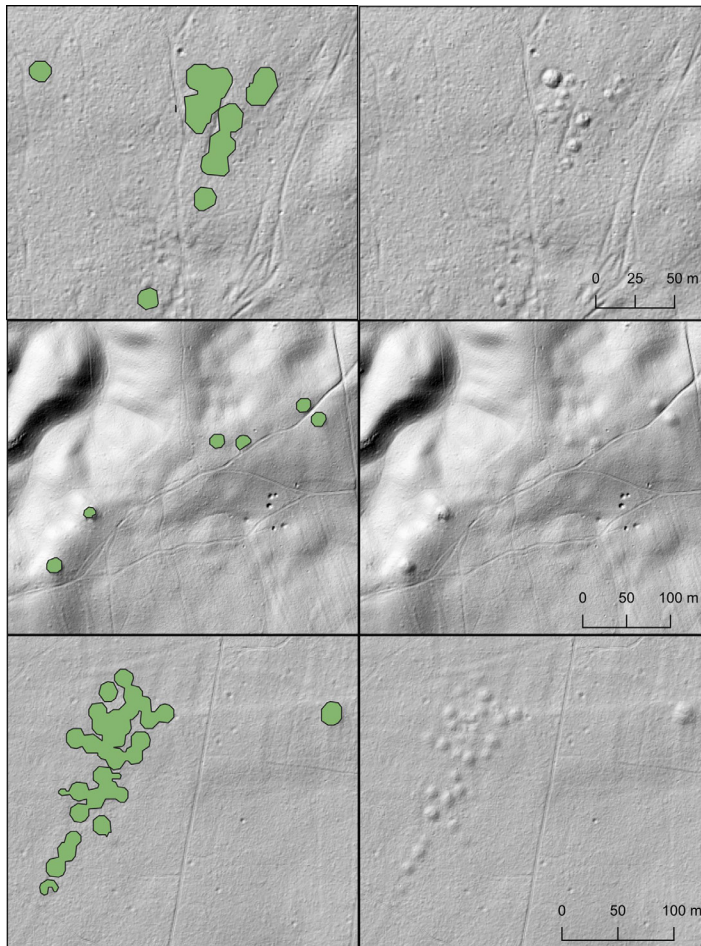
Inaczej niż przy konwencjonalnym podejściu w analizie z wykorzystaniem sieci neuronowych nie wyznaczyliśmy początkowo konkretnych obszarów testowych. Zamiast tego przyjęliśmy założenie, że jakość wyników będzie monitorowana i korygowana w sposób ciągły w miarę przeliczania kolejnych partii badanego obszaru. Proces polegał na stopniowym uczeniu i testowaniu modelu na coraz większym obszarze, przy jednoczesnej kontroli wyników i dostosowywaniu parametrów. Liczba obszarów wykorzystywanych do uczenia i walidacji była sukcesywnie zwiększana, a skuteczność działania detektora wzrastała równolegle. Proces ten był powtarzany aż do osiągnięcia obecnego poziomu dokładności.

Wynikiem prac są pliki wektorowe. Każdy wielokąt reprezentuje obszar, gdzie znajdują się kopce. Ponieważ sąsiadujące oznaczenia są łączone w jeden kształt, trudno jest jednoznacznie określić liczbę wykrytych obiektów. Sieć neuronowa wygenerowała dziewięć zestawów wyników, sklasyfikowanych od klasy 1 do klasy 9. Im wyższa klasa, tym wyższy poziom pewności modelu w przypisaniu obiektów do danej kategorii. Oznacza to, że w klasie 1 znajduje się najwięcej wielokątów, a w klasie 9 – najmniej. Takie podejście zapewnia użytkownikowi danych większą elastyczność w dostosowaniu wyników do własnych potrzeb i umożliwia efektywniejsze ich wykorzystanie w dalszej analizie.

Kontrola wyników działania sieci neuronowej

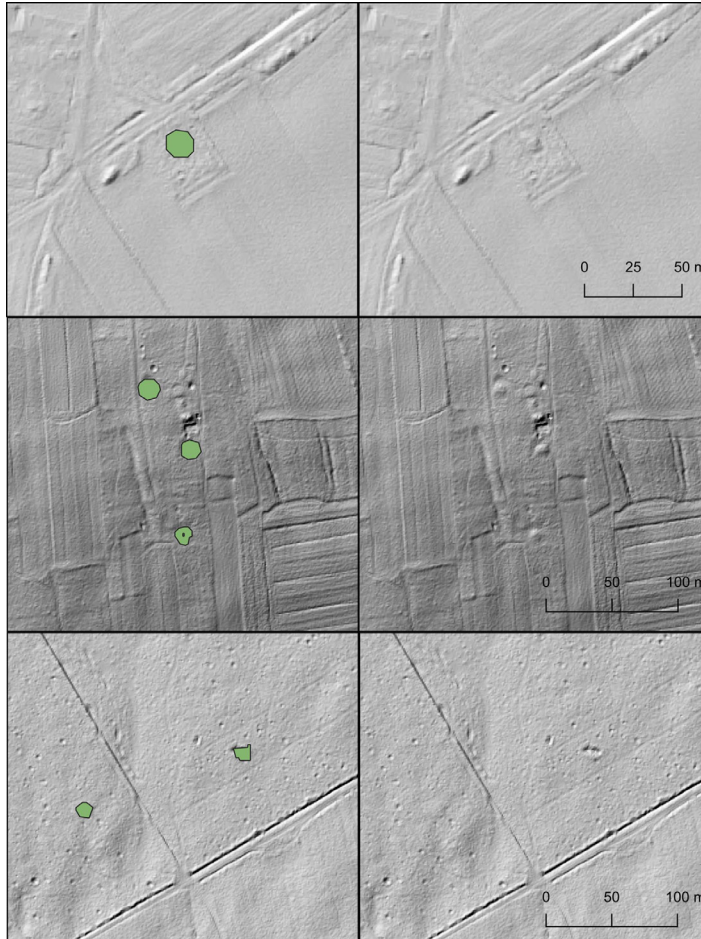
W uczeniu maszynowym miary takie jak precyzja (ang. *precision*), czułość (ang. *recall*) oraz miara F1 (*F1-score*) służą do oceny skuteczności modeli klasyfikacyjnych, szczególnie w razie występowania niezrównoważonego rozkładu klas. Możliwość obliczenia tych miar pozwala oprzeć klasyfikację na czterech podstawowych kategoriach wyników. Przypadki prawdziwie pozytywne (TP, od ang. *true positives*) w rzeczywistości należą do klasy pozytywnej i w rzeczywistości są kurhanami (ryc. 3). Wyniki fałszywie pozytywne (FP, od ang. *false positives*) oznaczają sytuacje, w których model błędnie zaklasyfikował jako pozytywne przypadki faktycznie należące do klasy negatywnej (ryc. 4). Prawdziwie negatywne (TN, od ang. *true negatives*) są przypadki należące do klasy negatywnej, które zostały poprawnie rozpoznane jako negatywne. Z kolei wyniki fałszywie

²⁴ T.-Y. Lin et al., *Focal Loss for Dense Object Detection* [w:] 2017 IEEE International Conference on Computer Vision (ICCV), [s.l.] 2017, doi.org/10.1109/ICCV.2017.324, dostęp: 18.05.2025.



3 Wizualizacja numerycznego modelu terenu i oznaczenia predykcji TP. Oznaczenia prawdziwie pozytywne w rzeczywistości pokrywają się z położeniem kurhanów

Visualization of the digital terrain model and true positive (TP) prediction markings. TP markings actually coincide with the location of burial mounds



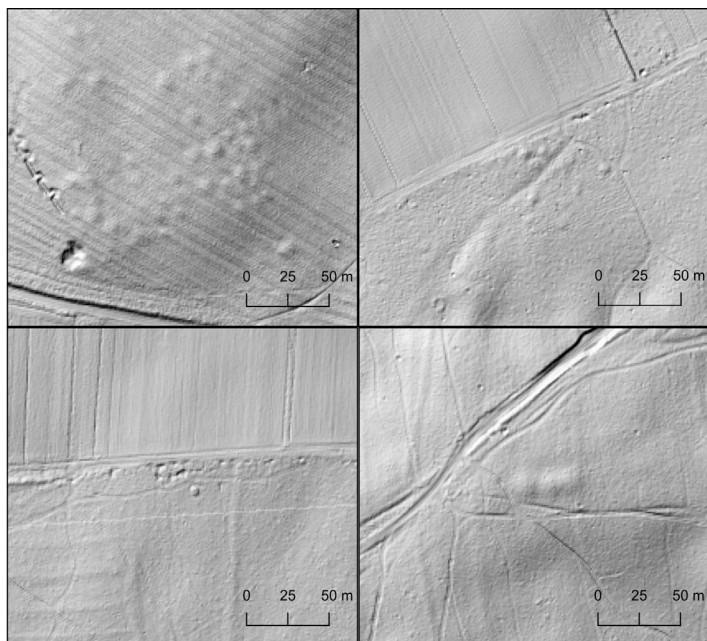
4 Wizualizacja numerycznego modelu terenu i oznaczenia predykcji FP. Oznaczenia fałszywie pozytywne to punkty, w których model błędnie zaklasyfikował jako pozytywne przypadki faktycznie niebędące kurhanami

Visualization of the digital terrain model and false positive (FP) prediction markings. FP markings are points where the model has incorrectly classified as positive cases that are not actually burial mounds

5

Wizualizacja numerycznego modelu terenu i oznaczenia predykcji FN. Brak oznaczeń to wynik fałszywie negatywny. Widać tu miejsca o morfologii typowej dla kurhanów lub ich skupisk, którym model nieprawidłowo przypisał klasę negatywną

Visualization of the digital terrain model and FN prediction markings. The absence of markings indicates a false negative result. Here, you can see places with a morphology typical of burial mounds, or clusters of burial mounds which the model has incorrectly assigned to the negative class



5

negatywne (FN, od ang. *false negatives*) oznaczają sytuacje, gdy model nieprawidłowo przypisał klasę negatywną przypadkom, które faktycznie należą do klasy pozytywnej (ryc. 5).

Precyzja określa odsetek poprawnie zaklasyfikowanych przypadków pozytywnych wśród wszystkich przypadków zaklasyfikowanych jako pozytywne, koncentrując się na trafności pozytywnych predykcji modelu. Oblicza się ją według wzoru:

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

Wyższa precyzja oznacza mniejszą liczbę FP.

Czułość (nazywana również wskaźnikiem trafień – ang. *true positive rate*) mierzy odsetek rzeczywistych przypadków pozytywnych prawidłowo rozpoznanych przez model, skupiając się na wykrywaniu jak największej liczby rzeczywistych pozytywów. Wzór na określenie czułości to:

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}$$

Wyższa wartość czułości sugeruje mniejszą liczbę fałszywych negatywów.

Miara F1 jest średnią harmoniczną precyzji oraz czułości i równoważy obie te miary. Oblicza się ją według wzoru:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

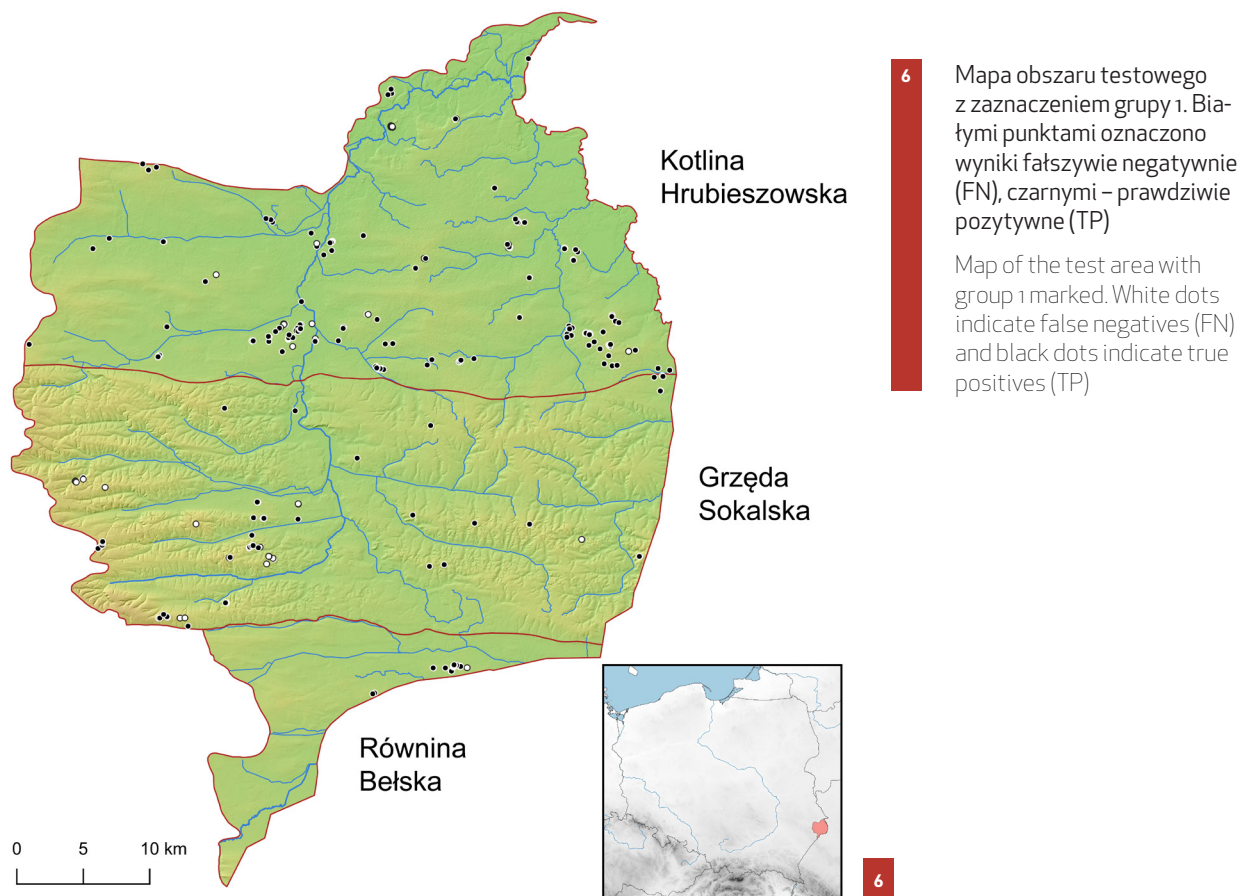
Miara F1 jest szczególnie przydatna przy niezrównoważonych danych, ponieważ uwzględnia zarówno fałszywe pozytywy, jak i fałszywe negatywy. Wysoka wartość tej miary świadczy o dobrym kompromisie między precyzją a czułością, co pozwala na bardziej kompleksową ocenę skuteczności modelu niż sama dokładność.

Wyniki

Na obecnym etapie projektu ocena jakości wyników stanowi pewne wyzwanie. Powodem są zarówno ograniczenia danych, jak i interpretacyjny charakter prospekcji archeologicznej. Nie mamy pewności, że wszystkie istniejące w obszarze testowym kopce zostały poprawnie zaanotowane ani że każda weryfikacja wyników modelu jest poprawna. W praktyce archeolog często mierzy się z pytaniem, czy obiekt widoczny w danych LiDAR rzeczywiście jest kurhanem, a metody weryfikacyjne nie zawsze pozwalają na jednoznaczną odpowiedź.

Obszar testowy, a więc wydzielony region niewykorzystany w trenowaniu i przeznaczony do niezależnej walidacji wyników modelu, obejmował Kotlinę Hrubieszowską, Grzędę Sokalską oraz Równinę Bełską o łącznej powierzchni około 1850 km². Na podstawie manualnej analizy danych LiDAR udało nam się zidentyfikować tam łącznie 489 kurhanów, które następnie podzielono na dwie grupy. Przypisanie do grupy zależało od stopnia pewności interpretacji. Grupa 1 obejmowała 384 „pewne” kurhany (ryc. 6), natomiast grupa 2 składała się ze 105 „niepewnych” obiektów (ryc. 7). Model sieci neuronowej dał wyniki zróżnicowane w zależności od grupy. Dla grupy 1 uzyskano 318 trafień (TP) i 66 pominięć (FN), co się przekłada na czułość równą 0,83. Dla grupy 2 odnotowano 47 trafień (TP) i 58 pominięć (FN), czułość wyniosła więc 0,45. Łącznie model wykrył 365 obiektów, pominął 124 i wygenerował 130 fałszywych trafień (FP) (ryc. 8), co dało precyzję i czułość na poziomie 0,74 ($F1 = 0,74$) (tab.). Wyniki te wskazują na dużą skuteczność detekcji obiektów jednoznacznych morfologicznie oraz na trudności w rozpoznawaniu form słabiej wykształconych lub ambiwalentnych.

Stała wartość $FP = 130$ raportowana przy obu grupach wymaga wyjaśnienia. Fałszywe trafienia definiujemy jako wszystkie predykcje modelu, które nie mają dopasowania do żadnego obiektu referencyjnego na obszarze testowym, niezależnie od tego, czy obiekt referencyjny należy do grupy 1 czy do grupy 2. Innymi słowy, FP tworzą jeden wspólny zbiór dla całego obszaru ewaluacji. Ponieważ obie grupy oceniano w tym samym przestrzennym zbiorze predykcji, przy



7 Mapa obszaru testowego z zaznaczeniem grupy 2. Białymi punktami oznaczono wyniki fałszywie negatywne (FN), czarnymi – prawdziwie pozytywne (TP)

Map of the test area with group 2 marked. White dots indicate false negative (FN) results and black dots indicate true positive (TP) results



7

8 Mapa obszaru testowego z zaznaczeniem wszystkich wyników fałszywie pozytywnych (FP)

Map of the test area with all false positive (FP) results marked



8

obliczaniu precyzji dla grupy 1 i grupy 2 użyto tej samej liczby FP. Dzięki temu uniknięto arbitralnego przypisywania fałszywych detekcji do jednej z grup. Wobec tego precyzji i F1 nie należy liczyć dla każdej grupy oddzielnie, a kluczową miarą różnic między grupami jest czułość (odsetek wykrytych obiektów w danej kategorii).

Tabela. Wyniki walidacji modelu detekcji kurhanów na obszarze testowym

	TP	FP	FN	Recall	Precision	F1
Grupa 1	318		66	0,83		
Grupa 2	47		58	0,45		
Suma	365	130	124	0,74	0,74	0,74

Należy zatem podkreślić, że w obecnej formie opracowana baza danych nie stanowi katalogu kurhanów. Niemniej jednak w naszym mniemaniu może być cennym źródłem danych do dalszych analiz. Badacz podejmujący się analizy danego regionu może wybrać jedną z dziewięciu dostępnych klas wyników, a tym samym dostosować dane w zależności od tego, czy preferuje wyższą czułość czy mniejszą liczbę fałszywych detekcji, bez konieczności pracy z siecią neuronową. Na obecnym etapie tworzymy więc bazę danych, a nie narzędzie, co jest o tyle użyteczne, że daje użytkownikowi gotowe dane, więc nie musi on wdrażać się w technologię.

Dyskusja

Główną przesłanką do sięgnięcia po metody uczenia maszynowego jest zwiększenie efektywności, skalowalności i powtarzalności analiz przy zachowaniu akceptowalnej jakości wyników. Decyzja o wdrożeniu takich metod nie powinna być jednak domyślna. Najlepiej, by była oparta na ocenie warunków brzegowych. Przed przystąpieniem do prac należy odpowiedzieć na kilka pytań.

- **Skala i czasochłonność:** czy ręczna analiza danego zbioru danych byłaby istotnie bardziej czasochłonna niż przygotowanie, trenowanie, walidacja i utrzymanie rozwiązania opartego na sieci neuronowej?
- **Bilans jakości:** czy zakładana oszczędność czasu nie będzie się wiązała z istotnym pogorszeniem jakości wyników, a jeśli tak, to czy wpływ tej utraty jakości na wnioski badawcze jest akceptowalny?
- **Dostępność i adekwatność danych:** czy dysponujemy wystarczająco dużym, reprezentatywnym i wiarygodnie zaanotowanym zbiorem danych zarówno do trenowania, jak i do niezależnej walidacji modelu?
- **Jednorodność obiektów:** czy wykrywane obiekty są wystarczająco spójne pod względem cech morfologicznych, aby umożliwić uogólnienie reguł decyzyjnych modelu poza obszar treningowy?

Pozytywna weryfikacja tych kryteriów uzasadnia wdrożenie metod uczenia maszynowego, w przeciwnym razie należy rozważyć podejście manualne lub hybrydowe.

W przypadku niewielkich zbiorów danych bardziej efektywne bywa ich tradycyjne opracowanie niż budowa i wdrażanie zautomatyzowanego rozwiązania. Wynika to z wysokich kosztów przygotowawczych: opracowania modelu, anotacji przykładów oraz testów walidacyjnych. To właśnie ta część stanowi wąskie gardło całego procesu, a drobne błędy popełnione na tym etapie mogą znacząco obniżyć skuteczność działania sieci. Kluczowe znaczenie mają zatem poprawna interpretacja danych oraz pełna, dokładna i konsekwentna anotacja prowadzona przez specjalistów z odpowiednimi kompetencjami. Dobrą praktyką jest podwójna, niezależna anotacja tego samego materiału, a następnie wspólna dyskusja nad rozbieżnościami. W naszym projekcie taka procedura wzajemnej weryfikacji zwiększyła liczbę wykrytych kopców w NMT o 10–15%.

Stworzenie efektywnego systemu wymaga szeroko zakrojonej współpracy interdyscyplinarnej, angażującej specjalistów z różnych dziedzin²⁵. Już na etapie przygotowawczym poświęciliśmy wiele czasu na doprecyzowanie celów, zakresu prac, struktury bazy danych i formatów plików. Mimo to pojawiały się trudności wynikające z różnic w doświadczeniu i podejściu specjalistów reprezentujących różne dziedziny. W naszym przypadku konieczne było powiązanie ze sobą kompetencji z zakresu archeologii, systemów informacji geograficznej (GIS), lotniczego skaningu laserowego oraz sieci neuronowych. Kluczową rolę odgrywało połączenie zaplecza teoretyczno-technicznego z umiejętnościami analityczno-interpretacyjnymi: wiedza archeologiczna warunkowała poprawną anotację obiektów, GIS umożliwiał sprawne operowanie danymi przestrzennymi, a kompetencje programistyczne zapewniały przygotowanie, trenowanie i prawidłowe działanie modeli. Ponieważ jedna osoba rzadko ma wszystkie wymagane umiejętności, niezbędne są tworzenie zespołów specjalistów i skoordynowanie ich prac. Szczególnego znaczenia nabiera wówczas precyzyjna komunikacja, zwłaszcza wobec różnic terminologicznych między dyscyplinami. Przykładowo termin *feature* w archeologii oznacza co innego niż w uczeniu maszynowym. Spójność metodologiczną i wysoką jakość wyników zapewnia dopiero precyzyjne zdefiniowanie pojęć oraz ujednoczenie procedur.

Jednym z istotnych wyzwań związanych z wykorzystaniem sztucznej inteligencji jest praca z danymi. Należy podkreślić, że metody uczenia maszynowego wymagają dużych zbiorów danych – w zasadzie im większy zbiór, tym większe szanse na skuteczność modelu. Zarządzanie takimi zbiorami oraz utrzymanie w nich porządku stanowi jednak spore wyzwanie, zarówno organizacyjne, jak i techniczne. Konieczne jest nie tylko posiadanie wystarczającej mocy obliczeniowej, lecz także zarezerwowanie znacznej ilości przestrzeni na dysku na przechowywanie danych. Równocześnie istotny jest swobodny i szybki dostęp do danych dla wszystkich uczestników projektu. Jednym z problemów, z którym zetknęliśmy się w toku prac, była wielkość bazy danych wysokościowych, która zawierała ponad 190 000 plików. Obsługa takiego wolumenu danych w środowisku QGIS okazała się w zasadzie niemożliwa – nawet samo otwarcie folderu z tymi plikami w systemie Windows sprawiało trudności. Przesłanie tak dużej liczby plików na popularne platformy do przechowywania danych w chmurze również okazało się niewykonalne. Powszechnie wykorzystywane oprogramowanie zwykle nie zakłada potrzeby pracy na tak dużych zbiorach. Problem ten można rozwiązać, dzieląc dane na mniejsze zestawy i pracując na nich sekwencyjnie. Takie podejście wymaga jednak dodatkowej uwagi lub automatyzacji działań. Potencjalnym rozwiązaniem mogłoby być również zastosowanie specjalistycznego oprogramowania lub wydajniejszego sprzętu, jednak wiąże się to z kolejnymi kosztami. W naszym przypadku odpowiedzią było wykorzystanie Pythona, który pozwala na względnie swobodne operowanie dużymi zbiorami danych. Nie jest to jednak rozwiązanie intuicyjne ani łatwe do wdrożenia przez użytkowników bez doświadczenia w programowaniu.

Kolejnym wyzwaniem jest ujednoczenie (normalizacja) danych. Standardowe techniki uczenia maszynowego zakładają, że dane wejściowe mają taki sam rozkład statystyczny co dane treningowe. Każde istotne odchylenie od niego (przesunięcie rozkładu) zwiększa ryzyko błędnej klasyfikacji. Problem dotyczy zarówno parametrów technicznych (między innymi rozdzielczości, liczby kanałów, sposobu filtracji), jak i różnic w morfologii terenu. Model wytrenowany na materiałach z nizin Europy Środkowej będzie działał słabiej w regionach górskich, gdzie kontekst topograficzny jest zupełnie inny. Dobry przykład mogą stanowić mielerze: w Polsce zwykle przyjmują postać niskich, płaskich kopców, podczas gdy w Norwegii najczęściej mają formę platform wciętych w zbocza. Mimo identycznej funkcji ich sygnatura morfologiczna jest odmienna, co należy uwzględnić już na etapie anotacji, doboru obszarów treningowych i walidacyjnych oraz ustalania procedur normalizacji i standaryzacji.

²⁵ Por. M. Olivier, W. Verschoof-van der Vaart, *Implementing State-of-the-Art Deep Learning Approaches for Archaeological Object Detection in Remotely-Sensed Data. The Results of Cross-Domain Collaboration*, „Journal of Computer Applications in Archaeology” 2021, vol. 4, iss. 1, doi.org/10.5334/jcaa.78, dostęp: 18.05.2025.

Sieci neuronowe są złożonymi modelami statystycznymi, których zachowanie wynika przede wszystkim z regularności zawartych w danych uczących, a jedynie w ograniczonym stopniu z jawnych decyzji projektowych. Ich zastosowanie wymaga więc utrzymania równowagi między wychwytywaniem istotnego sygnału a ignorowaniem przypadkowych korelacji. Dobrym przykładem wyzwania interpretacyjnego są kurhany i kopce graniczne: morfologicznie często nie do odróżnienia w NMT, różnią się niekiedy jedynie kontekstem przestrzennym (lokalizacja, układ). W takiej sytuacji możliwe są co najmniej trzy scenariusze:

- model nie znajdzie stabilnych cech rozróżniających i będzie klasyfikował obiekty w sposób zbliżony do losowego;
- model nauczy się artefaktu korelacyjnego, na przykład zacznie wiązać kurhany z terenami równinnymi, a kopce graniczne - z pagórkowatymi, mimo że nie jest to reguła merytoryczna;
- model poprawnie wykorzysta informacje o kontekście przestrzennym i osiągnie dużą trafność.

Ponieważ reguły decyzyjne sieci są trudne do rekonstrukcji, nawet bardzo dobre wyniki mogą wynikać z eksploatacji uprzedzeń obecnych w danych, a nie z właściwego wnioskowania. Dlatego należy starannie zaplanować zakres kontekstu dostępnego modelowi, odzwierciedlając ten, z którego korzysta człowiek, oraz projektować zbiory treningowe i walidacyjne tak, by minimalizować korelacje pozorne. W razie podejrzenia uprzedzeń warto rozszerzyć walidację o przykłady diagnostyczne (na przykład kopce graniczne na terenach równinnych), które umożliwią wykrycie tych uprzedzeń i korektę.

Sposobem na zwiększenie skuteczności procesu jest wygenerowanie nowego numerycznego modelu terenu na podstawie reklasyfikowanej chmury punktów, ponieważ udostępniany przez Główny Urząd Geodezji i Kartografii gotowy NMT nie do końca odpowiada potrzebom archeologii i w niektórych przypadkach klasyfikuje kurhany jako „niską roślinność”²⁶. Zwiększenie liczby punktów zaklasyfikowanych jako grunt oraz rozdzielczości NMT znacząco poprawiłoby wyniki teledetekcji. Niestety przetworzenie surowych danych LiDAR w skali całego kraju wykraczało poza dostępne zasoby obliczeniowe i ramy czasowe tego projektu. Świadomie przyjęty kompromis mógł miejscami obniżyć jakość teledetekcji, lecz umożliwił przygotowanie ogólnokrajowej warstwy potencjalnych kurhanów ze zdefiniowaną skalą wiarygodności. Co istotne, zgromadzone anotacje oraz modułowy przebieg dają możliwość stosunkowo łatwego udoskonalania systemu.

W kontekście polskiego wkładu w dyskusję należy nawiązać również do wcześniejszych doświadczeń z użyciem sieci U-Net do segmentacji wizualizacji DTM pochodzących z obszaru Puszczy Białowieskiej²⁷. Mimo stosunkowo niewielkiego zbioru uczącego, obejmującego 211 kurhanów, autorzy uzyskali wartościowe wyniki. Jakość z pewnością podwyższyło połączenie różnych sposobów wizualizacji danych i wprowadzenie w modelowanie klas konkurencyjnych. Jednocześnie odnotowano niedorozpoznanie małych kurhanów i drobne szумы. W naszym przypadku odpowiedzią na te problemy było znaczące zwiększenie liczby anotowanych obiektów. Wykorzystanie 1158 przykładów kopców dało znacząco lepsze rezultaty niż nasze wcześniejsze próby na mniejszych zbiorach.

Podsumowanie

Zastosowanie metod uczenia maszynowego to proces złożony, wymagający realizacji szeregu następujących po sobie zadań, aby osiągnąć oczekiwany rezultat (ryc. 2). Na obecnym etapie rozwoju tej technologii trudno wyobrazić sobie program, który działałby na zasadzie prostego wczytania danych wejściowych i szybkiego uzyskania wiarygodnych wyników. Jest to proces dynamiczny, który będzie nieustannie rozwijany i doskonalony. Wraz ze wzrostem jakości dostępnych danych

²⁶ G. Kiarszys, G. Szalast, *Archeologia w chmurze punktów. Porównanie rezultatów filtracji i klasyfikacji gruntu w projekcie ISOK z wynikami opracowanymi w oprogramowaniu LAStools i Terrasolid*, „Folia Praehistorica Posnaniensia” 2014, t. 19, doi.org/10.14746/fpp.2014.19.12, dostęp: 18.05.2025.

²⁷ P.Z. Banasiak et al., *Semantic Segmentation (U-Net) of Archaeological Features...*, op. cit.

pojawi się potrzeba dalszego poprawiania istniejących modeli, ponownej anotacji danych oraz testowania, a niekiedy także opracowywania nowych rozwiązań od podstaw. Równolegle konieczne jest rozwijanie odpowiednich procedur i podstawowych zasad, które będą nadawały kierunek rozwojowi projektów opartych na sztucznej inteligencji. W najbliższych latach planujemy zwiększyć nasz zbiór treningowy oraz wykorzystać nowsze architektury głębokich sieci neuronowych.

Potencjał, jaki ma ta technologia, stanowi istotne wyzwanie dla całej społeczności naukowej. W naszej ocenie sztuczna inteligencja nie stanowi jednak zagrożenia dla roli naukowca. Staje się natomiast niezwykle potężnym narzędziem, które w niedalekiej przyszłości będzie dla archeologa równie nieodzowne jak szpachelka.

Na obecnym etapie nie planujemy udostępnienia całej bazy danych w otwartym dostępie, głównie ze względu na ochronę stanowisk archeologicznych. Niemniej nasze wyniki w formie danych przestrzennych chętnie udostępniamy na prośbę badaczy lub przedstawicieli służb konserwatorskich.

Finansowanie

Niniejszy artykuł został przygotowany w ramach projektu finansowanego przez Narodowe Centrum Nauki pt. *Długie domy jako element złożonego krajobrazu kulturowego. Rekonstrukcja sieci osadniczej Kujaw w 2 poł. V tys. cal BC z użyciem najnowszych metod teledetekcyjnych* (nr 2019/35/B/HS3/02941).

Funding

This article was prepared as part of a project funded by the National Science Centre, entitled: *Długie domy jako element złożonego krajobrazu kulturowego. Rekonstrukcja sieci osadniczej Kujaw w 2 poł. V tys. cal BC z użyciem najnowszych metod teledetekcyjnych* (Long houses as an element of a complex cultural landscape. Reconstruction of the settlement network in Kujawy in the second half of the 5th millennium BCE using the latest remote sensing methods) (project no.: 2019/35/B/HS3/02941).

dr Michał Jakubczak

Archeolog zatrudniony w Laboratorium Bio- i Archeometrii Instytutu Archeologii i Etnologii Państwowej Akademii Nauk w Warszawie. Kierownik projektu *Długie domy jako element złożonego krajobrazu kulturowego. Rekonstrukcja sieci osadniczej Kujaw w 2 poł. V tys. cal BC z użyciem najnowszych metod teledetekcyjnych* (NCN 2019/35/B/HS3/02941). Specjalizuje się w teledetekcji archeologicznej, a także badaniach relikwów pradziejowych pól uprawnych.

Michał Jakubczak, PhD

Archaeologist employed at the Bio- and Archaeometry Laboratory of the Institute of Archaeology and Ethnology at the Polish Academy of Sciences in Warsaw. Manager of the project: *Długie domy jako element złożonego krajobrazu kulturowego. Rekonstrukcja sieci osadniczej Kujaw w 2 poł. V tys. cal BC z użyciem najnowszych metod teledetekcyjnych* (Long houses as an element of a complex cultural landscape. Reconstruction of the settlement network in Kujawy in the second half of the 5th millennium BC using the latest remote sensing methods) (project no.: 2019/35/B/HS3/02941). He specializes in archaeological remote sensing technologies and research into the relics of ancient farmland.

dr Michał Tyszkiewicz

Specjalista w zakresie informatyki i uczenia maszynowego. W 2023 roku obronił doktorat na Politechnice Federalnej w Lozannie (École polytechnique fédérale de Lausanne).

Michał Tyszkiewicz, PhD

He specializes in computer science and machine learning. In 2023, he obtained his doctorate from the Swiss Federal Institute of Technology in Lausanne (École polytechnique fédérale de Lausanne).

mgr Michał Leloch

Archeolog, doktorant w Szkole Doktorskiej Nauk Humanistycznych Uniwersytetu Warszawskiego. Stypendysta projektu *Wysokogórskie stanowiska ze sztuką naskalną w wąwozie Kyzyl Dara w Uzbekistanie* (NCN 2019/35/O/HS3/03051). Specjalizuje się w badaniach sztuki naskalnej.

Michał Leloch, MA

Archaeologist and doctoral student at the Doctoral School of Humanities at the University of Warsaw. He is a scholarship holder on the *Wysokogórskie stanowiska ze sztuką naskalną w wąwozie Kyzyl Dara w Uzbekistanie* (High-Altitude Sites with Rock Art in the Kyzyl Dara Gorge in Uzbekistan project (NCN 2019/35/O/HS3/03051). He specializes in research into rock art.

Bibliografia

- Abate Nicodemo et al., *An Open-Source Machine Learning-Based Methodological Approach for Processing High-Resolution UAS LiDAR Data in Archaeological Contexts. A Case Study from Epirus, Greece*, „Journal of Archaeological Method and Theory” 2025, vol. 32, doi.org/10.1007/s10816-025-09706-8, dostęp: 18.05.2025.
- Anglisano Anna et al., *Supervised Machine Learning Algorithms to Predict Provenance of Archaeological Pottery Fragments*, „Sustainability” 2022, vol. 14, iss. 18, doi.org/10.3390/su141811214, dostęp: 18.05.2025.
- Anichini Francesca et al., *The Automatic Recognition of Ceramics from Only One Photo. The ArchAIDE App*, „Journal of Archaeological Science. Reports” 2021, vol. 36, doi.org/10.1016/j.jasrep.2020.102788, dostęp: 18.05.2025.
- Argyrou Argyro, Agapiou Athos, *A Review of Artificial Intelligence and Remote Sensing for Archaeological Research*, „Remote Sensing” 2022, vol. 14, iss. 23, doi.org/10.3390/rs14236000, dostęp: 18.05.2025.
- Arnoldussen Stijn et al., *Field Systems and Later Prehistoric Land Use. New Insights Into Land Use Detectability and Palaeodemography in the Netherlands Through LiDAR, Automatic Detection and Traditional Field Data*, „Archaeological Prospection” 2023, vol. 30, iss. 3, doi.org/10.1002/arp.1891, dostęp: 18.05.2025.
- Banasiak Paweł Z. et al., *Semantic Segmentation (U-Net) of Archaeological Features in Airborne Laser Scanning – Example of the Białowieża Forest*, „Remote Sensing” 2022, vol. 14, iss. 4, doi.org/10.3390/rs14040995, dostęp: 18.05.2025.
- Berganzo-Besga Iban et al., *Hybrid MSRM-Based Deep Learning and Multitemporal Sentinel 2-Based Machine Learning Algorithm Detects Near 10k Archaeological Tumuli in North-Western Iberia*, „Remote Sensing” 2021, vol. 13, iss. 20, doi.org/10.3390/rs13204181, dostęp: 18.05.2025.
- de Boer Arjan, *Using Pattern Recognition to Search LIDAR Data for Archeological Sites [w:] The World is in Your Eyes. CAA2005. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 33rd Conference, Tomar, March 2005*, ed. Alexandra Figueiredo, Gonçalo Leite Velho, Tomar 2007, s. 245–254.
- Davis Dylan S., Sanger Matthew C., Lipo Carl P., *Automated Mound Detection Using Lidar and Object-Based Image Analysis in Beaufort County, South Carolina*, „Southeastern Archaeology” 2019, vol. 38, iss. 1, doi.org/10.1080/0734578X.2018.1482186, dostęp: 18.05.2025.
- Durham Peter, Lewis Paul, Shennan Stephen, *Artefact Matching and Retrieval Using Generalised Hough Transform [w:] Computer Applications and Quantitative Methods in Archaeology 1993*, ed. John Wilcock, Kris Lockyear, BAR International Series 598, [s.l.] 1995, s. 25–30.
- Evans Damian, *Airborne Laser Scanning as a Method for Exploring Long-Term Socio-Ecological Dynamics in Cambodia*, „Journal of Archaeological Science” 2016, vol. 74, doi.org/10.1016/j.jas.2016.05.009, dostęp: 18.05.2025.
- Freeland Travis et al., *Automated Feature Extraction for Prospection and Analysis of Monumental Earthworks from Aerial LiDAR in the Kingdom of Tonga*, „Journal of Archaeological Science” 2016, vol. 69, doi.org/10.1016/j.jas.2016.04.011, dostęp: 18.05.2025.
- Fadzly Nik et al., *A Deep Learning Framework for Bone Fragment Classification in Owl Pellets Using YOLOv12*, „Scientific Reports” 2025, vol. 15, doi.org/10.1038/s41598-025-15906-9, dostęp: 18.05.2025.
- Gattiglia Gabriele, *Managing Artificial Intelligence in Archeology. An Overview*, „Journal of Cultural Heritage” 2025, vol. 71, doi.org/10.1016/j.culher.2024.11.020, dostęp: 18.05.2025.
- Grilli Eleonora, Remondino Fabio, *Machine Learning Generalisation across Different 3D Architectural Heritage*, „ISPRS International Journal of Geo-Information” 2020, vol. 9, iss. 6, doi.org/10.3390/ijgi9060379, dostęp: 18.05.2025.
- Guyot Alexandre, Hubert-Moy Laurence, Lorho Thierry, *Detecting Neolithic Burial Mounds from LiDAR-Derived Elevation Data Using a Multi-Scale Approach and Machine Learning Techniques*, „Remote Sensing” 2018, vol. 10, iss. 2, doi.org/10.3390/rs10020225, dostęp: 18.05.2025.
- Hansen Richard D. et al., *LiDAR Analyses in the Contiguous Mirador-Calakmul Karst Basin, Guatemala. An Introduction to New Perspectives on Regional Early Maya Socioeconomic and Political Organization*, „Ancient Mesoamerica” 2023, vol. 34, iss. 3, doi.org/10.1017/S0956536122000244, dostęp: 18.05.2025.
- Hesse Ralf, *LiDAR-Derived Local Relief Models A New Tool for Archaeological Prospection*, „Archaeological Prospection” 2010, vol. 17, iss. 2, doi.org/10.1002/arp.374, dostęp: 18.05.2025.
- Iakubovskii Pavel, *Segmentation Models*, 2019, github.com/qubvel/segmentation_models, dostęp: 18.05.2025.
- Jamil Amirah H. et al., *A Review on Deep Learning Application for Detection of Archaeological Structures*, „Journal of Advanced Research in Applied Sciences and Engineering Technology” 2022, vol. 26, iss. 1, doi.org/10.37934/araset.26.1.714, dostęp: 18.05.2025.
- Karamitrou Alexandra et al., *Towards the Use of Artificial Intelligence Deep Learning Networks for Detection of Archaeological Sites*, „Surface Topography. Metrology and Properties” 2022, vol. 10, iss. 4, doi.org/10.1088/2051-672X/ac9492, dostęp: 18.05.2025.

- Kiarszys Grzegorz, Szalast Grzegorz, *Archeologia w chmurze punktów. Porównanie rezultatów filtracji i klasyfikacji gruntu w projekcie ISOK z wynikami opracowanymi w oprogramowaniu LAStools i Terrasolid*, „Folia Praehistorica Posnaniensia” 2014, t. 19, doi.org/10.14746/fpp.2014.19.12, dostęp: 18.05.2025.
- Kokalj Žiga, Hesse Ralf, *Airborne Laser Scanning Raster Data Visualization. A Guide to Good Practice*, Ljubljana 2017, doi.org/10.3986/9789612549848, dostęp: 18.05.2025.
- Kokalj Žiga, Somrak Maja, *Why Not a Single Image? Combining Visualizations to Facilitate Fieldwork and On-Screen Mapping*, „Remote Sensing” 2019, vol. 11, iss. 7, doi.org/10.3390/rs11070747, dostęp: 18.05.2025.
- Kokalj Žiga, Zakšek Klemen, Oštir Kristof, *Application of Sky-View Factor for the Visualisation of Historic Landscape Features in Lidar-Derived Relief Models*, „Antiquity” 2011, vol. 85, s. 263–273.
- Kurczyński Zdzisław, Stojek Edyta, Cisko-Lesicka Urszula, *Zadania GUGiK realizowane w ramach projektu ISOK [w:] Podręcznik dla uczestników szkoleń z wykorzystania produktów LiDAR*, red. Piotr Wężyk, Warszawa 2014, s. 22–58.
- Lin Tsung-Yi et al., *Focal Loss for Dense Object Detection* [w:] 2017 IEEE International Conference on Computer Vision (ICCV), [s.l.] 2017, doi.org/10.1109/ICCV.2017.324, dostęp: 18.05.2025.
- Maiza Chaouki, Gaildrat Véronique, *Automatic Classification of Archaeological Potsherds* [w:] *The Eight International Conference on Computer Graphics and Artificial Intelligence, 3IA'2005*, ed. Dimitri Pléménos, [s.l.] 2005.
- Miera Jan J. et al., *Large-Scale Investigations of Neolithic Settlement Dynamics in Central Germany Based on Machine Learning Analysis. A Case Study from the Weiße Elster River Catchment*, „PLOS One” 2022, vol. 17, iss. 4, doi.org/10.1371/journal.pone.0265835, dostęp: 18.05.2025.
- Olivier Martin, Verschoof-van der Vaart Wouter, *Implementing State-of-the-Art Deep Learning Approaches for Archaeological Object Detection in Remotely-Sensed Data. The Results of Cross-Domain Collaboration*, „Journal of Computer Applications in Archaeology” 2021, vol. 4, iss. 1, doi.org/10.5334/jcaa.78, dostęp: 18.05.2025.
- Orengo Hector A. et al., *Automated Detection of Archaeological Mounds Using Machine-Learning Classification of Multisensor and Multitemporal Satellite Data*, „PNAS” 2020, vol. 117, iss. 31, doi.org/10.1073/pnas.2005583117, dostęp: 18.05.2025.
- Paszke Adam et al., *PyTorch. An Imperative Style, High-Performance Deep Learning Library*, [w:] *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, ed. Hanna Wallach, [s.l.a.], s. 8024–8035.
- Piccoli Chiara et al., *Towards the Automatic Classification of Pottery Sherds. Two Complementary Approaches* [w:] *Across Space and Time. Papers from the 41st Conference on Computer Applications and Quantitative Methods in Archaeology (Perth, 25–28 March 2013)*, ed. Arianna Traviglia, Amsterdam 2015, s. 463–474.
- Quintus Seth, Davis Dylan S., Cochrane Ethan E., *Evaluating Mask R-CNN Models to Extract Terracing Across Oceanic High Islands. A Case Study from Samoa*, „Archaeological Prospection” 2023, vol. 30, iss. 4, doi.org/10.1002/arp.1909, dostęp: 18.05.2025.
- Reynolds Paul, *Trade Networks of the East, 3rd to 7th Centuries – the View from Beirut (Lebanon) and Butrint (Albania) (Fine Wares, Amphorae and Kitchen Wares)* [w:] *LRCW 3. Late Roman Coarse Wares, Cooking Wares and Amphorae in the Mediterranean. Archaeology and Archaeometry – Comparison Between Western and Eastern Mediterranean*, vol. 1, ed. Simonetta Menchelli et al., Oxford 2010, s. 89–114.
- Riley Melanie A., *Automated Detection of Prehistoric Conical Burial Mounds from LIDAR Bare Earth Digital Elevation Models*, Master's thesis, Department of Geology and Geography, North Missouri State, 2009.
- Ronneberger Olaf, Fischer Philipp, Brox Thomas, *U-Net – Convolutional Networks for Biomedical Image Segmentation* [w:] *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, „LNCS”, vol. 9351, ed. Nassir Navab et al., Cham 2015, doi.org/10.1007/978-3-319-24574-4_28, dostęp: 18.05.2025.
- Stott David, Kristiansen Søren M., Sindbæk Søren M., *Searching for Viking Age Fortresses with Automatic Landscape Classification and Feature Detection*, „Remote Sensing” 2019, vol. 11, iss. 16, doi.org/10.3390/rs11161881, dostęp: 18.05.2025.
- Sevara Christopher et al., *Pixel Versus Object – A Comparison of Strategies for the Semi-Automated Mapping of Archaeological Features Using Airborne Laser Scanning Data*, „Journal of Archaeological Science. Reports” 2016, vol. 5, doi.org/10.1016/j.jasrep.2015.12.023, dostęp: 18.05.2025.
- Stewart Adam J. et al., *TorchGeo. Deep Learning with Geospatial Data* [w:] *SIGSPATIAL '22. Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, New York 2022, doi.org/10.1145/3557915.3560953, dostęp: 18.05.2025.
- Toler-Franklin Corey et al., *Multi-Feature Matching of Fresco Fragments*, „ACM Transactions on Graphics (TOG)” 2010, vol. 29, iss. 6, doi.org/10.1145/1882261.1866207, dostęp: 18.05.2025.
- Zakšek Klemen, Oštir Kristof, Kokalj Žiga, *Sky-View Factor as a Relief Visualization Technique*, „Remote Sensing” 2011, vol. 3, iss. 2, doi.org/10.3390/rs3020398, dostęp: 18.05.2025.
- Zakirov Azamat et al., *Artificial Intelligence in Archaeological Geophysics. Testing ChatGPT-4o on Magnetometer Data from Sapallitepa (Uzbekistan)*, „Archaeological Prospection” 2025, doi.org/10.1002/arp.70006, dostęp: 18.05.2025.