ARCHAEOLOGICAL SURFACE CERAMICS DETECTION USING LOW ALTITUDE IMAGES BASED ON WEAK LEARNERS

Argyro Argyrou, Athos Agapiou

Department of Civil Engineering and Geomatics, Faculty of Engineering and Technology, Cyprus University of Technology, Saripolou 2-8, 3036 Achilleos 1 Building, 2nd Floor, Lemesos, Cyprus, P.O Box. 50329, 3603

ABSTRACT

Over the past decade, remote sensing sensors and their products have been increasingly utilized for archaeological science and cultural heritage studies. In our study, we explored the application of several supervised machine learning classifiers using red-green-blue (RGB) and multispectral high-resolution drone imagery to evaluate their performance towards semi-automatic surface ceramic detection. The results indicate that using low-altitude remote sensing sensors and advanced image-processing techniques can be incredibly innovative in archaeological research. However, our study also revealed existing research limitations in detecting surface ceramics, which significantly impact the detection accuracy. Therefore, detecting surface ceramics using RGB or multi-spectral drone imagery should be reconsidered as an 'imbalanced data distribution' problem. A new and robust methodology needed to be developed to address this "accuracy paradox" of imbalanced data samples and optimise archaeological surface ceramic detection. Our study aimed to fill a gap in the literature by blending AI methodologies for non-uniformly distributed classes.

Index Terms— Archaeology, Machine Learning (ML), Object detection, Classification, Earth Observation

1. INTRODUCTION

In recent years, remote sensing science has been progressively applied to support archaeological research [1,2]. Technological innovation and improvements of space-based sensors, in terms of their spatial and spectral resolution, and the adoption of open access and free distribution of satellite datasets (see Landsat and Sentinel products), have motivated the further use of space-based remote sensing applications [3]. In addition, the democratization of low-altitude systems, with drones at relevant low costs, has been widely adopted in the last decade in archaeological research, mainly for documentation purposes [4]. At the same time, archaeological computational approaches have shifted from desktop-based applications to cloud-based approaches blended with advanced AI algorithms [5]. As these changes have taken place in a relevant short period (a decade), one can argue that we are still in the beginning of a new era of so-called "remote sensing archaeology".

Orengo and Garcia-Molsosa in 2019 [5] demonstrated the great potential using the Google Earth Engine big data cloud platform, for detecting archaeological potsherds based on the acquisition of high-resolution UAVs red, green, blue (RGB) images and the application of random forest classification. Their work is pioneer as this was the very first example aiming to advance low altitude technologies with remote sensing methodologies for archaeological surface detection. Their work was based on a supervised Random Forest classification using training samples from the high-resolution orthophoto of their area of interest. Recently, in 2021, the authors also improved their previous work by developing a novel framework for detecting surface pottery, showing that low-altitude remote sensing sensors (e.g., Unmanned Aerial Vehicles, UAVs) can provide significant outcomes.

2. CASE STUDY

The purpose of this research is to explore how Artificial Intelligence (AI) algorithms can be used to classify and detect archaeological surface ceramics using low-altitude sensor cameras and how an interdisciplinary approach between Remote Sensing and Artificial intelligence can overcome the 'accuracy paradox' phenomenon of working with imbalanced remote sensing data.

When we classify data, we usually consider a model with higher accuracy the best. However, this may not be the case when dealing with imbalanced data. Imbalanced data sets have a disproportionate number of samples in different classes. When we use standard classifier learning algorithms to classify data with imbalanced class distribution, their performance can be greatly affected.

Various factors can affect the modeling of minority classes, including small sample size, separability, and the presence of sub-concepts within a class. To tackle this problem, this study employs an interdisciplinary methodology combining Remote Sensing with Artificial Intelligence. The goal is to address the paradoxical effect caused by these factors.

3. METHODOLOGY

This research methodology involves using Artificial Intelligence algorithms to classify archaeological ceramic surfaces. The images are captured through low-altitude sensor cameras. We used two drones to acquire drone-based images of the selected area of interest. Two flight campaigns were performed using the DJI Phantom 4 Pro system (spectral bands: Blue (B): 468 nm ± 47 nm; Green (G): 532 nm ± 58 nm; Red (R): 594 nm ± 32.5 nm), and the DJI P4 Multi-spectral system (spectral bands: Blue (B): 450 nm ± 16 nm; Green (G): 560 nm ± 16 nm; Red (R): 650 nm ± 16 nm; Red edge (RE): 730 nm ± 16 nm and Near-infrared (NIR): 840 nm ± 26 nm). The process is illustrated in diagram 1 below.

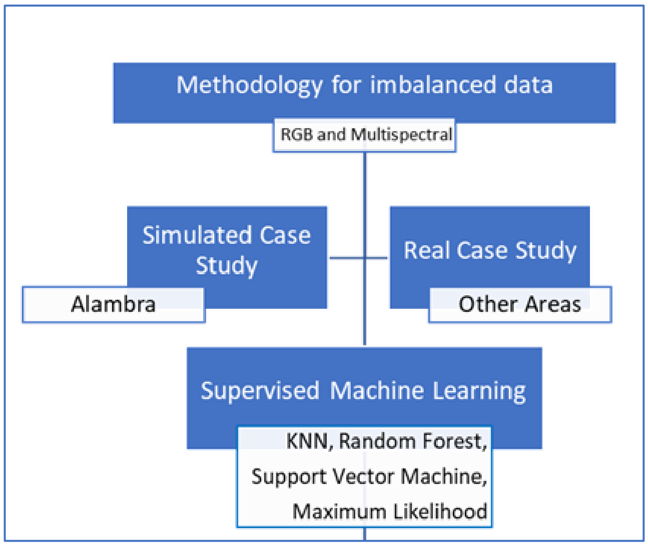

Diagram 1: Research methodology.

ArcGIS Pro software's Image Analyst tools were used for computational processing. One of the main challenges faced was dealing with imbalanced remote sensing data. To overcome this, the Training Samples Manager in the Classification Tools was used to create a training model comprising of three classes - 'ceramics' (class 1), 'soil' (class 2), and 'crops' (class 3). To create the initial training data, polygons were drawn on top of visible ceramic fragments, bare soil, and crops. The values of the pixels delimited by the polygons in each composite band were assigned to each class. Finally, four supervised classifiers - K-Nearest Neighbour (KNN), Random Forest (RF), Support Vector Machine (SVM), and the Maximum Likelihood algorithm were applied to the data. The maximum number of samples per class was set to 500 for the SVM, RF, and KNN classifiers, as this was deemed sufficient to ensure optimal results. The first step in the process was to classify the composite images using a trained classifier. Then, the classification was compared to the orthomosaic to assess its accuracy. To do this, randomly sampled points were created for post-classification accuracy assessment. The Accuracy Assessment Points tool from the Image Analyst tools evaluated all classification results. Randomly distributed samples were created for each class with an equal number of samples and then compared with the classification results. Based on the confusion matrix per classifier, the user's and producer's accuracy for each class and the overall kappa index were calculated. This procedure was implemented for RGB and multispectral drone images, and all results were extracted and evaluated on a local computer. Many standard classifier learning algorithms assume an equal distribution of classes and misclassification costs. However, these algorithms may not perform well when used on imbalanced data sets. An imbalanced data set is one where the class proportions are skewed, where the majority class has a larger proportion of samples than the minority class (in this case, archaeological surface ceramics).

In data analysis, it is often observed that the distribution of data in an imbalanced dataset is heavily skewed towards one class. This can pose a significant challenge while modelling minority classes, such as ceramics. Several factors can influence the modelling of such classes, including small sample sizes, separability, and the presence of sub-concepts. Small sample sizes lead to a lack of representation of the minority class, making it difficult to train a model that can accurately classify them. Separability refers to the degree to which the minority class is distinct from the majority class. If the minority class is poorly separable from the majority class, it can result in a higher misclassification rate. Lastly, sub-concepts within the minority class can pose a challenge, as different sub-concepts may require different models to be accurately classified.

Based on the results of our experiment, it appears that the issue stems from an imbalance between the surface ceramics and the surrounding environment, which includes the soil and crops. To dive deeper into this problem, we require more accurate classification results and effective tools to manage imbalanced data or update learning algorithms. The literature offers various solutions, such as rebalancing the class distribution by resampling the data space at the data level or modifying existing classifier learning algorithms to improve comprehension of the small class of ceramics at the algorithmic level. Binary classification problems can be challenging for boosting algorithms when dealing with imbalanced data. However, boosting algorithms are frequently utilized to enhance the prediction capabilities of weak learners and transform them into strong learners. Among the three primary boosting algorithms, AdaBoost, or Adaptive Boost, is commonly employed as an Ensemble Method in Machine Learning. The fundamental idea behind AdaBoost is to build a model and assign identical weights to all data points. It subsequently assigns higher weights to incorrectly classified points, and the following model emphasises those points with higher weights. This training process continues until a lower error rate is achieved. The basic steps of the research methodology are illustrated in diagram 2 below.

Diagram 2: Further analysis methodology and tools to treat imbalanced data.

4. PRELIMINARY RESULTS

The classifiers used for this study were trained using image samples for three classes: 'ceramics' (class 1), 'soil' (class 2), and 'crops' (class 3). The overall accuracy was estimated using randomly distributed testing pixels to evaluate their performance. Accuracy was calculated as the proportion of correctly predicted samples in the test set divided by the total predictions made on the test set.

Accuracy = Correct Predictions / Total Predictions

The accuracy for class 2 and class 3 (soil and crop) was estimated to be approximately 80%, while for class 1 (ceramics), a relatively low accuracy was reported for all four classifiers. The question was raised about how many testing pixels should be selected to ensure that the assessed accuracy was a reliable estimate of the actual accuracy. Would a larger sample of testing pixels give a more realistic estimate? What should the appropriate number of samples be?

According to [7], the number of samples required for an accuracy of 90% is 225, while 119 testing pixels are required for a 95% accuracy. These numbers assume that the classes follow a normal distribution, where a set of measurements, for instance, the mean, is distributed around the centre of these measurements.

Using 225 testing samples, the accuracy of all classifiers was estimated again. ArcGIS Pro randomly created 225 sampled points for post-classification accuracy assessment using the Image Analyst Toolbox. The sampling scheme was set to randomly distributed points, in which each class had the same number of points. A "Ground Truth" field and a "Classified" field were created in the final attribute table. Finally, we manually updated the Ground Truth field by changing or identifying the set of points and compared these fields using the Compute Confusion Matrix tool.

The ceramics class results showed a wide range of accuracy, with RGB images ranging from 12% to 24% and multispectral images ranging from 23% to 61%, as illustrated in Table 1 below. However, this presents a problem with the misclassification of minority classes. Classifiers tend to accurately predict the majority class while being ineffective in predicting the minority class. Our research proposes a methodology to detect surface ceramics using low-altitude multispectral and RGB cameras based on weak learners to address this issue.

Table 1: Ceramics Accuracy after supervised classification.

	KNN	MAX_Likelihood	SVM	RF
RGB	13%	12%	24%	15%
Multispectral	23%	52%	61%	31%

5. ACCURACY IMPROVEMENTS

Research shows that detecting surface ceramics using low-altitude sensors can yield significant results. However, there are limitations in accurately detecting the minority class of ceramics, which can impact the overall detection accuracy. To address this issue, we approach ceramic surface detection as an "imbalanced data distribution" problem. In previous studies, a problem with misclassifying minority classes, such as archaeological ceramics, was identified. Despite achieving high accuracy, the actual detection rate for the ceramic class remained low. This is because classifiers predict classes with extensive data more accurately than those with limited data.

5.1. Boosting algorithms

Boosting is a powerful machine-learning technique that harnesses the collective power of many rough rules of thumb to generate a single, highly accurate prediction rule. This is achieved by repeatedly using a weak learning algorithm on varying subsets of training examples until a new weak prediction rule is generated. After many iterations, the boosting algorithm combines these weak rules into a robust rule. Boosting is a technique that can be applied to any base learning algorithm by assigning greater weight to challenging examples and combining weak rules through majority voting.

In 1995, researchers [8] and [9] created the AdaBoost algorithm to improve the classification performance of other learning algorithms. AdaBoost is the first adaptive boosting algorithm that adjusts its parameters based on data performance. This includes re-weighting data and computing weights for final aggregation iteratively. The boosting algorithm comprises several steps, including:

Step 1: Initialise weights - At the beginning of the process, each training example is given equal weight.

Step 2: Train a weak learner - The weighted training data is used to train a weak learner. A weak learner is a simple model

that performs slightly better than random guessing. A decision tree with a few levels can be used as a weak learner.
Step 3: Error calculation - The error of the weak learner on the training data is computed. The weighted sum of incorrectly classified cases constitutes the error.
Step 4: Update weights - Weights are updated based on the error rate of the training examples. Misclassified examples are given higher weights, while correctly classified examples are given lower weights.
Step 5: Repeat - Steps 2 to 4 are repeated multiple times. A new weak learner is trained on each cycle's updated weights of the training examples.
Step 6: Combine weak learners - The final model comprises all the weak learners trained in the previous steps. The accuracy of each weak learner is weighted, and the final prediction is based on the weighted sum of the weak learners.
Step 7: Forecast - The completed model is used to predict the class labels of new instances.

Our decision-making framework now includes AdaBoost, resulting in significantly improved prediction accuracy compared to using only a decision tree. Ongoing interdisciplinary efforts involve testing and parameterizing algorithms, with the expectation of further improvements in accuracy.

6. CONCLUSION

Our research focused on determining the feasibility of using artificial intelligence techniques to automatically detect archaeological ceramics from high-resolution images captured by unmanned aerial vehicles (UAVs). We also aimed to establish a methodology that can deliver results in terms of time and accuracy comparable to, if not better than, those obtained through traditional archaeological field surveys. To achieve this, we employed supervised machine learning algorithms that utilized RGB and multispectral images acquired from a UAV.

Based on the methodology presented by [5], the findings of this study demonstrate that low-altitude remote sensing sensors can be highly effective in archaeological research. The classifiers employed in the study could predict majority classes (soil and crops) with high accuracy but could not accurately predict minority classes (surface ceramics). A new methodology has been proposed to improve the accuracy of detecting surface ceramics using drone images. This approach involves a boosting technique for weak learners and utilizes RGB and multispectral images. The resulting method provides more reliable and precise results, significantly improving the previous technique.

The authors plan to expand their research on ceramics to include different types from various periods and spectral behaviors in the same area. To achieve this, they will conduct lab-based spectral measurements to ensure statistically significant spectral separability among the ceramics during the same flight. The team scheduled drone surveys to improve data for algorithm training and outcome assessment. Aiming to reduce noise, enhance separability, and evaluate imbalanced ceramics data using measures like F-measure, G-mean, and ROC analysis.

11. REFERENCES

[1] A. Traviglia, A. Torsello, “Landscape pattern detection in archaeological remote sensing” Geosciences, 2017, 7, 128. [doi:10.3390/geosciences7040128]

[2] D. Tapete, “Remote Sensing and Geosciences for Archaeology” Reprinted from: Geosciences, 2018, 8(2), 41. [doi: 10.3390/geosciences8020041]

[3] R. Lasaponara, N. Masini, “Satellite remote sensing in archaeology: Past, present and future perspectives” Journal of Archaeological Science, 2011, 38(9), 1995-2002 [DOI: 10.1016/j.jas.2011.02.002]

[4] F. Materazzi, M. Pacifici, “Archaeological crop marks detection through drone multispectral remote sensing and vegetation indices: A new approach tested on the Italian pre-Roman city of Veii” Journal of Archaeological Science: Reports 41, 2022, 103235. [https://doi.org/10.1016/j.jasrep.2021.103235]

[5] H.A. Orengo, & A. Garcia-Molsosa, “A brave new world for archaeological survey: Automated machine learning-based potsherd detection using high-resolution drone imagery” Journal of Archaeological Science, 2019, 112, 105013. [https://doi.org/10.1016/j.jas.2019.105013]

[6] Y. Sun, A.K.C. Wong, M.S. Kamel, “Classification of imbalanced data: a review”, *International Journal of Pattern Recognition and Artificial Intelligence* 23(04), 2009, DOI: 10.1142/S0218001409007326, Sourcedx.doi.org

[7] J.A. Richards, *Remote Sensing Digital Image Analysis* (2022), Springer-Verlag Berlin Heidelberg, ISBN, 978-3-030-82327-6 (eBook), https://doi.org/10.1007/978-3-030-82327-6

[8] R.E Schapire, “The Strength of Weak Learnability”. *Machine Learning*. 1990;5(2):197{227.

[9] Y. Freund, “Boosting a Weak Learning Algorithm by Majority”. *In: Fulk MA, Case J, editors. Proceedings of the Third Annual Workshop on Computational Learning Theory*, COLT 1990, University of Rochester, Rochester, NY, USA, August 6-8, 1990; 1990. p. 202{216.

12. ACKNOWLEDGEMENT

This project has received funding from the European Union’s Horizon Europe Framework Programme (HORIZON-WIDERA-2021-ACCESS-03, Twinning Call) under the grant agreement No 101079377 and the UKRI under project number 10050486. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the UKRI. Neither the European Union nor the UKRI can be held responsible for them.