CS476 – Senior Capstone Design in Computer Science

Project Description

Project Title: ML Toolbox for Archaeology Applications	
Sponsor Information: 	Dr. Kayeleigh Sharp, Ph.D., RPA, Assistant Teaching Professor, Archaeology Director, Digital Archaeology Lab, NAU Director, Northern Gallinazo Archaeology Project, Lambayeque Department of Anthropology Northern Arizona University kayeleigh.sharp@nau.edu

Project Overview:

This project aims to overcome the basic weakness of handheld x ray fluorescence spectrometer (pXRF) analysis through the creation of an ML tool to predict the location of metal and pottery production (and other areas of anthropogenic activity). While it is highly sensitive and capable of detecting trace composition of many chemical elements, the spatial extent of its analysis is highly restricted (ca. 5 mm in diameter). The challenge is combining the analytical capability of pXRF (small scale) and GIS data (large scale) using AI or ML tools. We propose that by combining pXRF with macro-scale archaeological survey we can develop a high-resolution, evidence-based understanding of activities carried out in the past that moves beyond more traditional methods of intra-site archaeological detection. This is important as human activities leave chemical traces in soils that may be otherwise invisible to scientists studying the past.

At Huaca Letrada on the north coast of Peru, we aim to investigate the spatial and social contexts of certain activities, primarily copper procurement, metalworking and pottery production (i.e., smelting, firing and manufacture) using: (1) smartphone and UAV (drone) site recording and sfM (structure from motion) 3D image analysis, (2) GPS mapping, and (3) analysis of soils, architecture and artifacts with pXRF. Implementation of this multipronged approach makes it possible to elicit a refined understating of differential use of anthropogenic spaces. The disadvantage of this approach is that such datasets, although relevant and informative, are extremely difficult to bring together in a single stage of analysis.

The novel combination of data that we propose has important implications for resolving uncertainties related to variable use of functional spaces, inability to detect these with the human eye which often results in costly and time-consuming exploratory excavation which damages sites and is often unproductive.

By advancing non-destructive, ML guided surveys, we aim to standardize an extremely multi-scale approach to data analysis and interpretation in archaeology. As a member of this team, you will design an ML tool that will help to streamline prediction of production loci to help target areas of interest for further investigation, to identify loci that may be disrupted by human activities such as farming or landscaping and preserve others for upcoming generations. As with all prediction algorithms, your tool will quantify the uncertainty of its predictions as well.

When completed, the ML tool will improve ability for systematic archaeological site detection and quantitative approaches to data integration. It will also provide a means for analyzing archaeological sites that is consistent with

the needs of stakeholders (property owners or descendant communities, researchers, and governmental agencies) while at the same time expanding understanding of the past.

The solution we need (our mvp):

- An app interface for data collection
- A standard data packaging process
- An AI or ML tool that uses imagery + pXRF + GPS data for classification of sherds (and other soil/mining-derived anthropogenic items) to predict GPS locations for sources of clay/material/processing.

The solution components that would be really cool (stretch goals):

- the ability to share this ML tool with archeologists around the world
- the ability to deploy this tool on non-specialized equipment (smartphones or tablets) in the field for real-time site assessment.

Knowledge, skills, and expertise required for this project:

- Programming & ML techniques
- Some familiarity with 2D and 3D terrain modeling (i.e., sfM or photogrammetry).
- Basic understanding of aerial image analysis and processing

Equipment Requirements:

- There should be no equipment or software required other than a development platform and software/tools freely available online.

Software and other Deliverables:

- A strong as-built report detailing the design and implementation of the product in a complete, clear and professional manner. This document should provide a strong basis for future development of the product.
- Complete professionally documented codebase delivered both as a repository in GitHub, BitBucket, or some other version control repository; and as a physical archive on a USB drive.
- Software Deliverables:
    - An ML tool to integrate multiple data types and predict archaeological activity areas.
    - A well-tested, cross-platform app that facilitates collection archeological data (may be more than imagery, pXRF, and GPS data).
    - Stretch goal: A central, secure data storage and service solution that will support the collection of archeological data from many researchers globally. This solution should allow the serving of data for the development and use of ML models within the broader archeological community.