Ex Machina: Machine Learning, ceramics and rock art in the Khorat Plateau, Thailand

Ex Machina: แมชชีนเลิร์นนิง เซรามิกส์ และศิลปะหินบนที่ราบสูงโคราช ประเทศไทย

Felise Goldfinch
College of Arts, Society and Education, Division of Tropical Environments and Societies,
James Cook University
Felise.Goldfinch@my.jcu.edu.au

Dr Nigel Chang
College of Arts, Society and Education, Division of Tropical Environments and Societies,
James Cook University
Nigel.Chang@jcu.edu.au

Dr Dianna Hardy
Indigenous Education and Research Centre,
James Cook University
Dianna.Hardy@jcu.edu.au

ABSTRACT

Machine Learning for the recognition and analysis of prehistoric rock art and pottery is a promising area of research that could reveal new insights into cultural heritage and identity. Deep Learning (a form of Artificial Intelligence) can now be used to train powerful models to automatically recognise pottery and rock art images, overcoming resource constraints such as time, manpower, and lack of funding. This article provides a preliminary overview and proof of concept by providing Machine Learning approaches based on current advancements in Deep Learning to train a model to recognise images of pottery and prehistoric rock art. These methods can process large amounts of data quickly and accurately, revealing new patterns and relationships. Although ML can be a complex undertaking, new tools make it accessible to the archaeological practitioner who is not an AI expert.

แมชชีนเลิร์นนิงสำหรับการจดจำและวิเคราะห์ศิลปะหินยุคก่อนประวัติศาสตร์และเครื่องปั้นดินเผาเป็นพื้นที่ที่มีแนวโน้มของการวิจัยที่สามารถเปิดเผยข้อมูลเชิงลึกใหม่ๆ เกี่ยวกับมรดกทางวัฒนธรรมและเอกลักษณ์ ตอนนี้สามารถใช้การเรียนรู้เชิงลึก (ปัญญาประดิษฐ์รูปแบบหนึ่ง) เพื่อฝึกโมเดลที่ทรงพลังให้จดจำภาพศิลปะเครื่องปั้นดินเผาและหินโดยอัตโนมัติ เอาชนะข้อจำกัดด้านทรัพยากร เช่น เวลา กำลังคน และการขาดเงินทุน บทความนี้แสดงภาพรวมเบื้องต้นและการพิสูจน์แนวคิดโดยให้แนวทางการเรียนรู้ของเครื่องตามความก้าวหน้าในปัจจุบันของการเรียนรู้เชิงลึกเพื่อฝึกโมเดลให้จดจำภาพของเครื่องปั้นดินเผาและศิลปะบนหินยุคก่อนประวัติศาสตร์ วิธีการเหล่านี้สามารถประมวลผลข้อมูลจำนวนมากได้อย่างรวดเร็วและแม่นยำ เผยให้เห็นรูปแบบและความสัมพันธ์ใหม่ๆ แม้ว่า ML จะเป็นงานที่ซับซ้อน แต่เครื่องมือใหม่ๆ ทำให้ผู้ปฏิบัติงานทางโบราณคดีที่ไม่ใช่ผู้เชี่ยวชาญด้าน AI สามารถเข้าถึงได้

Keywords: Machine Learning, Deep Learning, Cultural Heritage, Ceramics, Rock art, Archaeology, Convolutional Neural Networks | การเรียนรู้ของเครื่อง, การเรียนรู้อย่างลึกซึ้ง, มรดกทางวัฒนธรรม, เซรามิกส์, ศิลปะหิน, โบราณคดี, เครือข่ายประสาท Convolutional

INTRODUCTION

The significance of rock art is recognised worldwide (Whitley 2016). However, a lack of resources often hampers the discovery and exploration of this unique kind of cultural heritage. Manually identifying rock art and analysing it can be a time-consuming and expensive endeavour. Automation of this process would allow larger-scale analysis, enabling this study to be extended to locations where there has been little previous research. Similar issues exist in the analysis of ceramics. In the last few years, advances in Machine Learning (ML), a subfield of Artificial Intelligence (AI), have begun to be applied to the study of ancient ceramics and rock art. Currently, much of this research is in the 'proof of concept' stage; nevertheless, it shows promise, particularly in the area of image classification (Pawlowicz et al. 2017; Prasomphan and Jung 2017; Tsigkas et al. 2020; Bickler 2021). In recent years, ML-based algorithms have been used to identify and extract unique qualities from a wide range of creative and cultural assets, including rock art. However, studies that report on such research assume familiarity with advanced data science and ML algorithms, making it difficult for archaeologists to learn and use these techniques.

Ban Non Wat (BNW) and Khao Chan Ngam (KCN), both of which are located on the Khorat Plateau in Thailand, are home to impressive collections of ceramics and rock art, although much of the data regarding these sites remain relatively unexamined. This paper discusses the application of deep learning techniques (DLs) using Convolutional Neural Networks (CNNs) to analyse ceramics and rock art image datasets from the Khorat Plateau and shows the effectiveness of these methods.
Three primary contributions are made in this paper:

- We outline the process needed to perform classification of images from two different sites containing very different types of material culture (rock art and ceramics) using DL
- We provide details regarding our use of lower code frameworks for image classification and a discussion of the benefits and difficulties associated with such solutions for archaeologists and other researchers who do not have access to data science and machine learning specialists.
- And finally, we provide a discussion of some of the implications surrounding the use of ML for image analysis and the limitations of methods using open-source and commercially hosted ML options.

This paper first presents a background on the topic of DL and how this may be used in archaeological research. Next, we describe the methods we use to perform image classification using CNNs. We propose a method of using transfer learning (TL), which makes use of a model previously trained on a very large image dataset (which is not made up of archaeological data) to analyse two image datasets. We then turn to a discussion of the implications of using DL; both from a general ML standpoint and from the perspective of archaeology. Finally, we discuss our conclusions and recommendations for future work.

BACKGROUND

Machine Learning and Deep Learning Application in archaeology

ML seeks to answer the question of how to build machines with the capacity to learn independently (Jordan and Mitchell 2015). Computer models are digital representations of the physical world. A weather model may calculate how likely it is to rain given a particular set of parameters (called features). After the model has been trained by ingesting a large number of examples of previous outcomes, it is capable of responding with a prediction of how likely a particular outcome is to occur. Over time, the model is capable of 'learning', that is, making a more accurate prediction of the outcome is likely to occur as it collects more example data. Models are essentially a collection of simple calculations (addition, multiplication, and logarithmic functions) that are executed at great speed.

Deep Learning (DL) is a sub-field of ML and is concerned with creating "neural networks" that mimic the structure of networks of individual human brain cells. The concept of neural nets has been around since 1960 (Widrow and Lehr 1990), but only recently have advances in processing and computing power made them practical. Researchers have developed complex networks with many layers of neural nets that can execute formulas with hundreds of thousands of parameters. They are run on systems with graphic processing units (GPUs) or tensor processing units (a tensor is a type of mathematical vector) to allow these calculations to be executed extremely quickly. DL models are capable of analysing text to look for patterns of words, images of elements in a photo, or sounds such as the human voice. Many businesses are using some type of DL in their customer interactions (Davenport and Mittal 2023).

Convolutional Neural Networks (CNNs) are a type of layered artificial neural network used for image and video recognition tasks and are the engine that powers image classification. They are designed to process and analyse data with grid-like structures, such as images, videos, and audio signals. The convolutional layers are responsible for learning and detecting local patterns and features in the input data, while the pooling layers reduce the spatial dimensions of the data while preserving important information. The fully connected layers are used to make the final prediction or classification. The key advantage of CNNs is their ability to automatically learn hierarchical representations of the input data, without the need for manual feature engineering. They can also be trained on large amounts of labelled data, which allows them to learn and generalise well to new, unseen data. CNNs have been successfully applied in a variety of computer vision tasks, including object recognition, image classification, and semantic segmentation, among others (El Naqa and Murphy 2015).

There is great promise in applying AI techniques to the study of rock art. However, in the archaeological domain, visual item categorisation, one of an archaeologist's primary techniques, has been mostly unassisted by technology. Although visual inspection is traditionally used to classify artefacts, ML techniques can be used to construct tools to assist in this task (Maaten et al. 2007). Literature describing the application of ML to painted rock art analysis is limited; however, there have been a few notable publications in this area, such as Purshotam's (2015) master's thesis on automatic indexing of South African rock art images, Kowlessar et. al.'s (2021) model that identifies a stylistic chronology from learned features, and Jalandoni et al.'s (2022) article on the use of ML methods in rock art research and their application to automatic painted rock art identification.

Stages of the ML workflow

The stages involved in using ML for artefact classification are summarised in Table 1 below. These activities as they relate to CV and image analysis will be discussed in more detail in the methods section.

	Stage	Activities
0	Data collection	Collect a representative and diverse dataset of artefact images or other relevant information for training and testing the ML model.
1	Data pre-processing	Clean and pre-process the data to remove errors, missing values, or inconsistencies and ensure that the data are in a suitable format for ML.
2	Feature extraction	Extract meaningful features from the data that are relevant to artefact classification. This can include information such as the shape, size, colour, texture, and other characteristics of the artefacts.
3	Model selection	Choose a suitable ML algorithm for artefact classification. This may involve testing different algorithms and comparing their performance on the data.
4	Training and validation	Train the ML model on the data using a suitable training set and validate its performance on a separate validation set.
5	Model evaluation	Evaluate the performance of the trained model using appropriate metrics, such as accuracy, precision, and recall and identify any areas for improvement.
6	Model deployment*	Deploy the trained model for real-world artefact classification tasks, and use it to make predictions about new, unseen artefact data.
7	Model refinement*	Continuously evaluate and refine the model as new data become available, and retrain the model as needed to improve its performance.

Table 1 Stages in the ML workflow. *These stages are beyond the scope of the report of this paper.

The details and implementation of these stages may vary depending on the nature of the artefact data, the complexity of the classification task, and the availability of relevant resources and tools. Of these stages, data pre-processing requires the bulk of time needed to make data usable by an automated model, and so we turn our attention next to a discussion of the steps involved in this process.

Data pre-processing

Pre-processing as applied to ML is the action of ensuring that the data are in a suitable format and condition to be used by a model. Regardless of the domain where ML is used, pre-processing consumes most of the time required to use ML. A standard rule of thumb is that 80% of the time for a ML project will be consumed by data pre-processing (Frye et al. 2021). The next two sections are concerned with the steps in the pre-processing stage and their importance to the archaeological process and then a discussion of the similarities and differences between data pre-processing and Principal Component Analysis.

Pre-processing image data for use in ML

Data pre-processing consists of preparing and transforming raw data into a format that can be used as input to ML algorithms. This stage is particularly important when the data is obtained from different sources such as different sites or different orientation of the camera angle to the artefact in the image. It involves tasks such as cleaning and transforming data, handling missing values, and normalising data. Processing archaeological data is a time-consuming task that often falls to graduate students or entry-level employees. Data pre-processing is not the most glamorous work, but it is essential to ensure that the data are accurate and usable by the ML algorithm. The risk is that if processing data is seen as a low-priority task, it may be rushed or done incorrectly.

The steps in data pre-processing include:

- Data cleaning: Identify and remove any errors or inconsistencies in the data. Actions: detect and remove images that are too dissimilar, handle missing values such as no labels, and deal with duplicate data.
- Data Transformation: Convert data to a suitable format for ML algorithms. This process can include converting text data into numerical data or transforming images into arrays of pixel values suitable for the ML model.
- Data normalization: Adjusting the data to ensure that it is on the same scale and has the same distribution. Datasets with unequal distribution are skewed and may produce inaccurate predictions. Add or remove data as needed (i.e., choose more images of a certain type). ML models assume that data are suitably scaled and are proportional.
- Data augmentation: Add copies of the data that are modified such as tilted in angle, flipped, mirrored, zoomed, cropped, etc. For the best accuracy, DL requires large datasets with many examples. Many DL frameworks provide methods to automate this process. Overfitting occurs when a model learns the detail and noise in the training data to such an extent that it has a negative impact on the model's performance when applied to new data.
-

Image Classification Methods

This section describes the image categorisation approach we propose for ceramics and rock art. The study area where the data was collected will be first presented, followed by the DL models utilised. For various DL models, images are often scaled to a lower resolution to reduce the number of parameters to be learnt. We apply a DL approach using transfer learning (TL) with a model that had been previously trained on a large set of non-archaeological images. The following are details about the setup for the two experiments and what methods were followed to execute the machine learning code. We include a description of a free development environment, Google Colab, which we used to conduct the experiments, as well as input data dimensions. Lastly, we discuss performance evaluation measures and the generalisability of the models.

Study areas

The Khorat Plateau (ที่ราบสูงโคราช) takes its name from the local dialect of Nakhon Ratchasima Province and is located in northeastern Thailand incorporating a number of provinces, including parts of Nakhon Ratchasima, Buriram, and Maha Sarakham (Figure 1). The plateau is surrounded by uplands and is characterized by fertile soil, abundant water resources, and a dry- tropical climate. Two basins are separated from the plateau by the Phu Phan Mountains: the northern Sakhon Nakhon Basin and the southern Khorat Basin. The region is drained by several rivers, including the Mun and Chi rivers, which flow into the Mekong River forming the region's northeastern boundary. The geography of the Khorat Plateau has played a significant role in shaping its cultural and economic development, historically making it a barrier that controlled access to the region. This made it an important center for agriculture, trade, and transportation in the region. Historically, this made the plateau difficult to access. After the Post-Angkor period (Keyes 1976) and a protracted series of droughts between the 13th and 15th centuries, the plateau appears to have been mostly depopulated. The Khorat Plateau in Northeast Thailand has significant archaeological importance due to its rich cultural heritage and abundant ancient remains. The plateau has been occupied by various civilizations and cultures over the centuries and has played an important role in the development of the region. Archaeological sites on the plateau, such as Ban Chiang and BNW, reveal early examples of rice cultivation, metalworking, and advanced social organisation. The Khorat Plateau is therefore considered a crucial area for the study of Southeast Asian prehistory and the development of early civilizations in the region. Keyes (1974: 504) explains:

> "… to acquire a fuller and more accurate picture of the society and culture of the early urban life on the Khorat Plateau, much more archaeological investigation must be conducted. And, I would add, we can gain a deeper knowledge of these civilisations by undertaking rigorous research on a number of the Thai-Lao people of north-east Thailand's indigenous myths."

Forty years later Higham (2014) adds, "...we remain essentially ignorant of the linkages between sites and the presence or absence of states on the Khorat plateau". The plateau is considered a crucial area for the study of Southeast Asian prehistory and the development of early civilizations in the region, with artefacts found there providing important insights into the lifestyles, beliefs, and technological advancements of the civilizations that inhabited the area. Elaborating on this Higham and Kim (2022: 592) suggest that "sociocultural complexity during the local Iron Age was multidimensional and multilinear… [however] complex polities- early state analogues are still out of focus in the archaeological record". Despite a wealth of archaeological evidence, much remains unknown about the society and culture of early urban life on the Khorat Plateau, and more research is needed to gain a deeper understanding of these civilizations. The datasets used for the DL analysis reported in this paper come from two sites in the Khorat Plateau: BNW and KCN. Background information regarding these two sites is provided next.

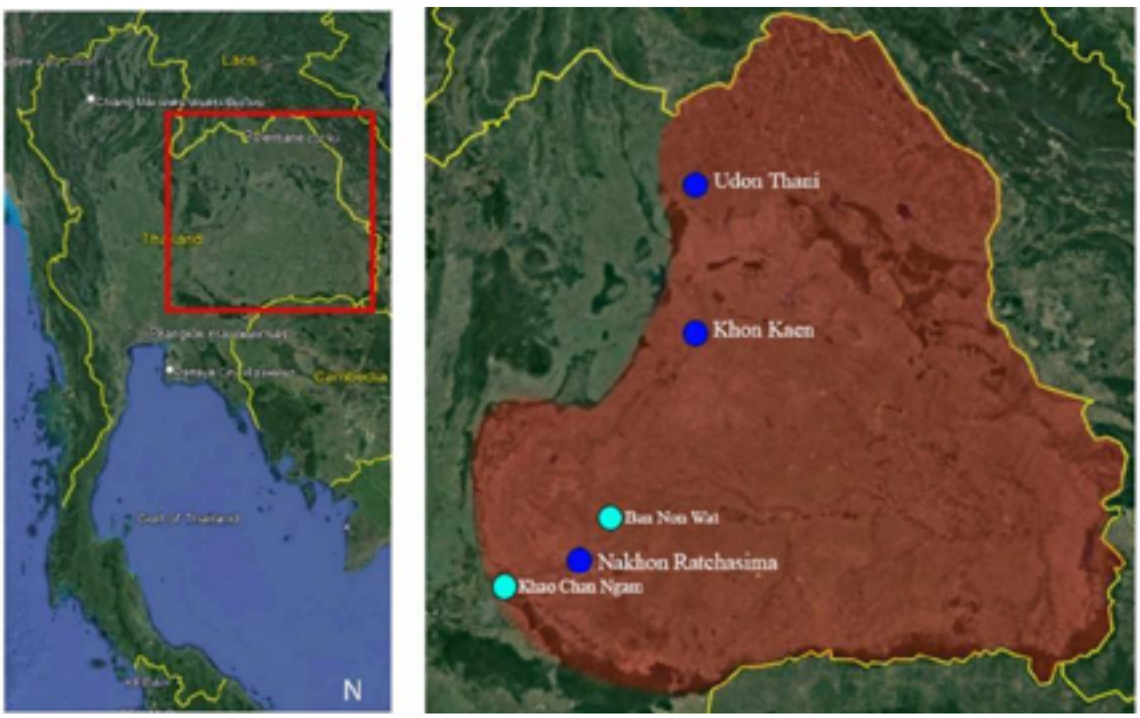

Fig. 1 Map of Thailand showcasing the Khorat Plateau and the locations of the study sites (after Google Earth Pro 2023).

Ban Non Wat
BNW is an archaeological site located in northeastern Thailand on the Khorat Plateau. It is a Thai town in the Non-Sung district of Nakhon Ratchasima province, close to the small city of Phimai. The region was occupied during the Neolithic, Bronze, and Iron Ages, according to archaeological evidence, and provides insights into the development of early civilisations in Southeast Asia (Harris and Tayles 2012). It has been the subject of excavations since 2002, revealing a cultural sequence that spans 11 prehistoric periods and 640 burials (Higham 2011). BNW is particularly significant, as it is believed to have been a key centre of early bronze production and trade. The first Neolithic settlement in BNW occurred in the 17th century BCE, whereas the Bronze Age began in the late 11th century BCE. The transition to the Iron Age occurred around 420 BC (Higham 2011). Seventy-six radiocarbon readings and Bayesian computations have refined the exclusiveness of this sequence (Higham 2011; Higham and Higham 2009). Bayesian analysis is the application of Bayesian statistical interpretation of probability to radiocarbon dating to obtain a more exact date (Otárola-Castillo and Torquato 2018).

Khao Chan Ngam
KCN is located 45 kilometres from the provincial capital of the Nakhon Ratchasima Sikhio district and is connected to Ban Loet Sawat. KCN is situated within a large collection of sandstone ridges that provide a natural rock shelter at the base of the Petchabun mountain range. According to Khanthakan (1979) and Tan (2014), the sandstone shelter is the largest in the region. The monks who constructed Wat KCN in 1968 allow access to KCN through the forest temple. KCN is notable for its prehistoric paintings of humans and animals. There are two parallel, north-south oriented sandstone massifs at this location. On the side walls of the caves are pictures of people (men, women, and children) indulging in various activities, such as sitting, dancing, standing with a dog, and firing arrows with bows, which portray daily life between approximately 4,000 and 2,500 BP (Tan 2014).

Development environment and frameworks used
We selected Google Colab and Jupyter notebooks to conduct our machine learning experiments. Jupyter notebooks are self-contained files capable of executing complex scripts written in programming languages such as Python, R, and Scala. Google Colab is a free cloud service that enables the execution of Jupyter notebook code and provides Google Drive storage access. The combination of these two free tools allows researchers to conduct ML and DL experiments without installing costly software. Colab also offers access to advanced computer resources, such as graphics processing units (GPUs) and tensor processing units (TPUs), which are required to execute the complex calculations required by ML/DL quickly. For our experiments, we used PyTorch, an open-source ML library. It offers a comprehensive platform for building and training machine learning models, as for well as executing existing models. To reduce development time, we also used Fast.AI, a wrapper library that sits on top of PyTorch and simplifies the creation of ML and DL models.

Experimental workflow
To replicate the most common scenario for dataset acquisition, we chose to use two disparate archaeological image collections for our experiments. The ceramics dataset consisted of 450 photographs with easily identified ceramics (in other words, one pot per image) from the site of BNW. The dataset of rock art images consisted of 165 photographs with easily identified rock art from the site of KCN with one motif present in each image. Our goal was to use ML/DL to do binary classification on each image.

Data pre-processing

Because our datasets contained relatively few images, we expanded the datasets by performing data augmentation on the images. All images were resized to 224 x 224 pixels x 3 colour channels. Additionally, Fast.AI was used to automate image duplication to expand our datasets using image flips, image rotations, image squish, random resized crops, and padding the image images with zeros. For accuracy DL algorithms require many images to train on to “teach” the model how to generalise from training images which are labelled, to test images which are not.

Feature extraction

Feature extraction is the transformation of raw data into numerical features that may be handled while maintaining the integrity of the original data set. It yields better results than applying ML directly to raw data. ML on images is effective because features may be used to compare photographs and correlate with another (owing to similarity) or with a specific label. People can easily recognise a vehicle or a tree in an image. Even if you have never seen a particular tree or automobile before, you can accurately associate it with the relevant thing or compare it to other objects in your memory that are comparable. In the case of an automobile, the presence of wheels, doors, a steering wheel, etc., distinguishes a fresh instance from others. It occurs because you sense forms and elements outside the image itself; hence, you can distinguish a unique tree or car if it exhibits particular traits. Google’s “I am not a robot” CAPTCHA method is an example of feature extraction in practice. Benefits of ML include improved accuracy, efficiency, scalability, and the ability to automatically learn and improve from experience without being explicitly programmed; and the ability to deal with increased complexity as more data are used.

Model selection:

This case study employs two types of machine learning: Supervised, using Fast.ai and TL, using PyTorch. Both models used the RESNET18 Architecture and were compared for accuracy, speed, and ease of use. When using DL with a small number of examples models can overfit. A solution is to use a model that has been pre-trained on thousands of images.

Training and validation

Train-test split is a technique used in ML to evaluate the performance of a model. It involves dividing the original dataset into two parts: a training set and a testing set. The training set is used to train the model, and the model is fit to the data in this set. The testing set is then used to evaluate the performance of the model, by providing it with input data and comparing its predictions with the actual target values. The objective is to estimate the model’s performance on new, unseen data. This study used an 80:20 ratio for each dataset (with 132 rock art images for training: 33 rock art images for testing and with 350 ceramics images for training: 150 ceramic images for testing). The split was done randomly to ensure that no bias was introduced into the results. The train-test split is an important technique for evaluating the performance of machine learning models, as it allows us to estimate how well the model will perform on new data.

For these experiments binary manual labelling was used. Manual labelling is the most accurate method, but it can be time-consuming and expensive. For the binary classification in this study, the goal is to predict one of two possible outcomes for a given input, ‘figurative/non-figurative’, or ‘restricted/unrestricted’. Binary classification problems can be solved relatively quickly and this can be useful for gaining a preliminary understanding of the problem and testing different approaches to the problem. Binary classification problems often require smaller datasets, making it easier to train models and perform experiments with limited computational resources. As this study includes two separate image classification notebooks, each data set was divided into 'figurative' and 'non-figurative' for rock art and 'restricted' and 'unrestricted' for ceramics.

Results

In ML, the number of epochs used during training can impact the results of the model. Epochs define the number of times the model will see and learn from the training data. Generally, increasing the number of epochs will result in a more accurate model as the model has had more opportunities to learn from the data. However, increasing the number of epochs will also result in a longer training time.

Ban Non Wat Ceramics

Ceramics images from the BNW archaeological site were utilised to construct both a Supervised, or labelled, code and Transfer Learning code to train a model. This model was based on Cameron (2013) in which she devised a typology for a sample of reconstructed Bronze Age and Iron Age burial containers. In conjunction with seriation, Cameron's (2013) typology revealed significant variations in ceramic traditions from the beginning of the Bronze Age 2 (about 1250 BC) and the beginning of the Bronze Age 5 (approximately 1000 BC) (c. 690 BC). The initial typology consists of two broad categories: restricted (Figure 3a) and unrestricted (Figure 3b), with an additional 16 unique subclassifications that represented whether the degree to which a vessel was restricted or unrestricted. This is most evident in the ceramic' necks were open. However, for the initial image detection, the coding for this model was condensed into binary categories: restricted and unrestricted.

Fig. 3a: example of restricted ceramic (Chang, 2019)

Fig. 3b: example of unrestricted ceramic (Chang, 2019)

Fast.ai

Table 2 shows the results for the ceramic Fast.ai training process. Each row represents the results for a single epoch, or iteration, of the training process. This model used 4 epoch iterations and total time it took for this process can be estimated based on the time column in the table. For each epoch, the time taken to run the model is specified in the format "hh:mm:ss". Based on the time column, each epoch took approximately 1 minute and 43 seconds to run, so the total time for the four epochs was approximately 6 minutes and 52 seconds. Additionally, the results indicate that the train_loss and valid_loss decrease as the number of epochs increases, indicating that the model is becoming more accurate and better able to fit the data. Additionally, the error_rate decreases as the number of epochs increases, which is a further indication of the model's improved accuracy.

epoch	**train_loss**	**valid_loss**	**error_rate**	**time**
0	1.312665	1.415879	0.466667	01:43
epoch	**train_loss**	**valid_loss**	**error_rate**	**time**
0	0.929468	0.747595	0.355556	01:46
1	0.847627	0.362395	0.177778	01:43
2	0.652359	0.318893	0.155556	01:46
3	0.543068	0.311377	0.155556	01:43

Table 2 Fast.ai metrics error rate showing a decrease in inaccurate predictions.

Based on the findings (Table 2; Figure 4), it is evident that this code was successful. The systems were using the most basic classification of restricted vs unrestricted without the subsections in Cameron's typology with an approximate accuracy of 75%. When evaluating any of these systems, it is essential to keep in mind the expectations regarding the algorithm's level of accuracy. The definition of 'sufficient precision' in ML is highly subjective, with industry standards recommending a range of 70 to 90%. A confusion matrix is a mechanism for summarising the effectiveness of a classification system. If the dataset is balanced and you are using two classes, then classification accuracy should be reliable. By calculating a confusion matrix, you could gain a better understanding of what your classification model does right and what kinds of errors it makes. The confusion matrix (Figure 6) has rows for each 'restricted' and 'unrestricted' ceramic in the dataset. In general, the classifier does not produce many errors, and the data is only slightly overfitted.

Fig. 4 ML results for the Fast.ai categorisation of BNW ceramics with 'X' indicating an inaccurate prediction.

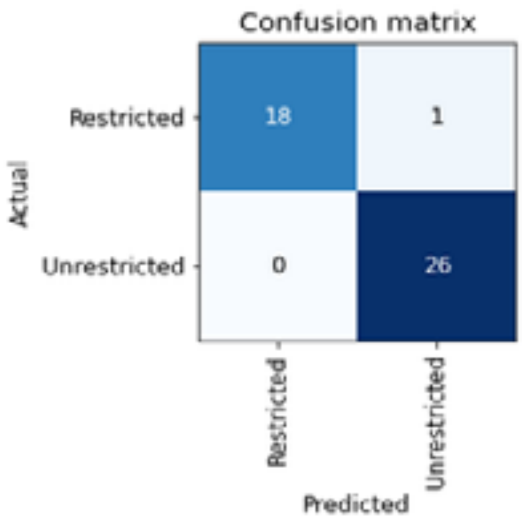

Fig. 5 Results for the Fast.ai BNW confusion matrix.

Transfer Learning

This approach trains a convolutional neural network for image classification using transfer learning. The model underwent two training and validation iterations, in which the model is being trained for 25 epochs (iterations) in each. For each epoch, the performance is evaluated on two sets of data: a training set and a test set. The "Loss" and "Acc" values represent the model's loss (error) and accuracy respectively.

Epoch 22/24

TRAIN Loss: 0.2960 Acc: 0.8811
TEST Loss: 0.8324 Acc: 0.6906

Epoch 23/24

TRAIN Loss: 0.2246 Acc: 0.9251
TEST Loss: 0.8086 Acc: 0.7130

Epoch 24/24

TRAIN Loss: 0.2285 Acc: 0.8811
TEST Loss: 0.9239 Acc: 0.6726

Table 3 Sample of Transfer Learning metrics error rate for Training and Validation 1.

Epoch 22/24

TRAIN Loss: 0.3796 Acc: 0.8370
TEST Loss: 0.7086 Acc: 0.6816

Epoch 23/24

TRAIN Loss: 0.4154 Acc: 0.7885
TEST Loss: 0.7551 Acc: 0.6592

Epoch 24/24

TRAIN Loss: 0.3450 Acc: 0.8590
TEST Loss: 0.7541 Acc: 0.6547

Table 4 Sample of Transfer Learning metrics error rate for Training and Validation 2.

The loss values in Table 3 show that the model's training accuracy improves over the course of training, but its test accuracy fluctuates. In this case, the model seems to be improving during the first few epochs, with a decrease in loss and an increase in accuracy on both the training and test datasets. However, the performance plateaus after a few epochs and starts to fluctuate. The model appears to reach a maximum test accuracy of around 71% in the final epoch. This first training and validation took approximately 1 hour 12 minutes 18 seconds. The results in Table 4 for Training and Validation 2 indicate that the model's training loss decreases and accuracy increases over time, reaching a maximum accuracy of 86.78% on the 15th epoch. Training loss and accuracy are improving with each epoch; however, test loss and accuracy oscillate at times. This is a common occurrence known as overfitting, where the model is performing well on the training data, but not generalising well to new data. The aim is to find a model that has a good balance between training performance and test performance. Additionally, the accuracy varies between different epochs, suggesting that the model may not be fully converging. The average runtime for Training and Validation two was 1 hour, 9 minutes and 56 seconds.

Khao Chan Ngam rock art

Analysis of rock art often involves attempting to identify the subject matter and interpreting the meaning behind the images. Photographs from Khao Chan were utilised to construct a Supervised code and a Transfer Learning code to train a model. This dataset of rock art images were based on an iconographic classification and were divided into two categories: Figurative (Figure 4a) and Non-Figurative (Figure 4b). Figurative rock art refers to images that depict recognisable objects, people, animals, or symbols.

Non-figurative rock art, on the other hand, consists of abstract designs or patterns that do not represent anything specific. One characteristic of this dataset is its skewness, as the rock art motifs at KCN are primarily figurative.

Fig. 4a Example of Figurative rock art (Tan, 2014).

Fig. 4b Example of Non Figurative rock art (Tan, 2014).

Fast.ai

Table 5 shows the results of the rock art Fast.ai training process. Like the Ceramics data, each row represents the results for a single epoch, or iteration, of the training process. Again, this model used 4 epoch iterations and total time it took for this process can be estimated based on the time column in the table. The total time for this process is approximately total of 00:29 seconds to train over the given epochs. Based on Table 5, the model improved in terms of both "train_loss" and "valid_loss" over the course of training and had a low error rate of 0.045455 in the validation dataset after each epoch.

epoch	train_loss	valid_loss	error_rate	time
0	1.247491	0.261175	0.045455	00:09
epoch	**train_loss**	**valid_loss**	**error_rate**	**time**
0	1.297137	0.199935	0.045455	00:06
1	1.248356	0.163435	0.045455	00:04
2	1.262464	0.180043	0.045455	00:04
3	1.131350	0.233866	0.090909	00:06

Table 5 Fast.ai metrics error rate.

The "error rate" column in the table represents the percentage of incorrect predictions made by the model. In Table 5, the error rate ranges from 4.55% to 9.09%. Therefore, the model had an accuracy of around 95.45%, which means that it correctly predicted the outcomes for approximately 95.45% of the data it was evaluated on. However, as shown, there are photographs that have been misinterpreted (Figure 5). Ordinarily, it is difficult to determine why an algorithm identifies photos as belonging to a particular category; interestingly, the two of the figurative images used were sitting, indicating that the system lacked sufficient training on those specific image types. Due to the few samples (images) in the dataset, the confusion matrix indicates that the model was unable to generalise knowledge to accurately predict rock art motifs (Figure 6).

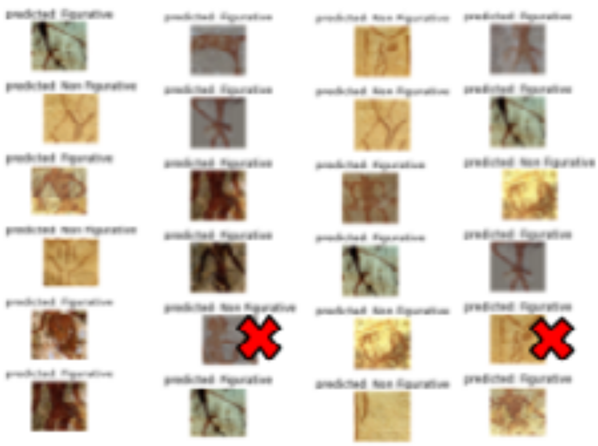

Fig. 5 ML results for Fast.ai categorization of KCN rock art.

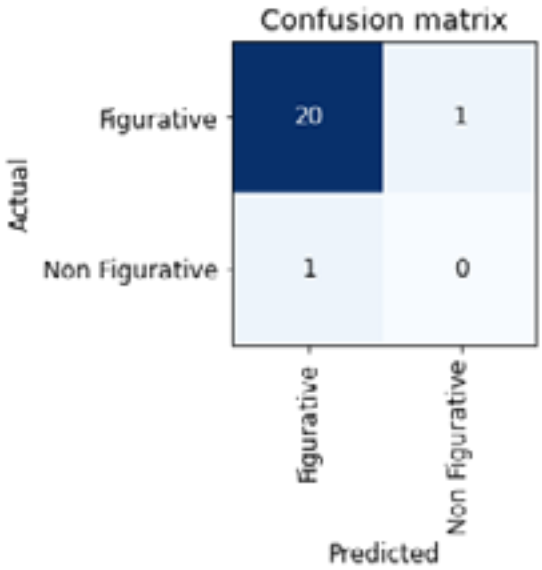

Fig. 6 Results for the Fast.ai KCN confusion matrix.

Transfer Learning

This model underwent two training and validation iterations in which the model was trained for 25 epochs on each. For each epoch, the performance is evaluated on two sets of data: a training set and a test set. The Training and Validation 1 results models train loss decreases over time and the train accuracy increases over time, indicating that the model is learning and improving on the training data. The test accuracy fluctuates but generally remains above 90%, with the best accuracy being 96.22% indicating that the model has a high level of accuracy. This first training and validation took approximately 2 minutes and 37 seconds.

Epoch 22/24

TRAIN Loss: 0.3885 Acc: 0.8326
TEST Loss: 0.6044 Acc: 0.7399

Epoch 23/24

TRAIN Loss: 0.3623 Acc: 0.8458
TEST Loss: 0.6621 Acc: 0.7220

Epoch 24/24

TRAIN Loss: 0.3893 Acc: 0.8018
TEST Loss: 0.6541 Acc: 0.7085

Table 6 Sample of Transfer Learning metrics error rate for Training and Validation 1.

Epoch 22/24

Train Loss: 0.0635 Acc: 0.9821
Test Loss: 0.1178 Acc: 0.9434

Epoch 23/24

Train Loss: 0.1592 Acc: 0.9732
Test Loss: 0.1284 Acc: 0.9434

Epoch 24/24

Train Loss: 0.0469 Acc: 0.9911
Test Loss: 0.1168 Acc: 0.9434

Table 7 Sample of Transfer Learning metrics error rate for Training and Validation 2.

Table 7 results indicate that the train loss decreases, and the train accuracy increases over the 24 epochs. Training loss and accuracy metrics are improving throughout the epochs, reaching a minimum of 0.0469 for the training loss and 1.0 for the training accuracy in epoch 16. In contrast, the test loss and accuracy metrics are fluctuating, reaching a maximum test accuracy of 0.9811 in epoch 5 and a minimum of 0.8491 in epoch 3. The test loss also decreases and the test accuracy increases, but the model may have overfitted to the training data as the gap between train and test performance increases as the number of epochs increases. It is important to note that a model with high accuracy on the training dataset does not always translate to high accuracy on the test dataset, as it might be overfitting to the training data. The overall accuracy of this model can be estimated to be 94.34% with an approximate run time of 2 minutes and 49 seconds.

DISCUSSION

ML and archaeology

This paper has provided results of two experiments using small collections of images of ceramics and rock art from the Khorat Plateau. Each dataset was trained and tested using the same algorithm (ResNet18) and following a supervised learning and transfer learning process. These experiments have demonstrated the following.

- That the small number of example images in a dataset can be problematic when using DL, however this limitation can be overcome by either using automated data augmentation to increase the number of example images or by using a model trained on thousands of images that are unrelated to archaeology.
- Demonstrated options for using lower code tools such as Jupyter notebooks, Google Colab and ML frameworks such as Fast.ai and PyTorch
- Demonstrated that existing code for doing CNN and TL is easily generalisable to disparate datasets of archaeological images without extensive revisions.
- Shown that some of the tedious work of classifying images into broad categories can be automated, leaving time for other explorations.

ML has the potential to revolutionise the field of archaeology by providing new tools and methods for analysing and processing large and complex data sets. This can improve both intersite and intrasite analysis by finding patterns and relationships in data that traditional methods cannot. Clustering algorithms can group sites and artefacts based on similarities, and decision tree models can predict cultural affiliations and the function of features. Ways ML has the potential to improve the speed and accuracy of archaeological research and help preserve cultural heritage include:

Data analysis: ML algorithms can be used to analyse large datasets of archaeological information, such as artefact and site data, and identify patterns and relationships that would be difficult to detect using traditional methods. For example, clustering algorithms can be used to group sites based on similarities in their artefact assemblages, while decision tree models can be used to predict the function of a feature within a site based on its location and other characteristics.

- Predictive modelling: ML algorithms can be used to make predictions about archaeological sites and artefacts, such as their cultural affiliation, age, or function. For example, decision tree models can be used to predict the cultural affiliation of a site based on its location and other characteristics, while neural networks can be used to predict the age of an artefact based on its physical characteristics.

- Cultural Heritage Management: ML can be used to support heritage management and preservation by identifying threats, predicting the impact of climate change, and monitoring the condition of cultural heritage sites.
- Artefact classification: ML can be used to classify and categorise artefacts, allowing more efficient and accurate cataloguing and management of archaeological collections.
- Image and signal processing: ML algorithms can be used to process and analyse images and signals from remote sensing and other imaging technologies, such as LiDAR and GPR. For example, convolutional neural networks can be used to detect and classify archaeological features in drone or satellite images, while signal processing algorithms can be used to analyse GPR data to identify subsurface features.
- Data visualisation: ML algorithms can be used to create visualisations of archaeological data, making it easier to understand and communicate complex relationships and patterns. For example, dimensionality reduction algorithms can be used to create 2D or 3D visualisations of artefact and site data, while clustering algorithms can be used to visualise patterns and relationships in the data.

General implications

Limitations of Google Colab

Google Colab offers the opportunity to create Jupyter Notebooks to run ML code on the fly; however, it is not without issues. Access to Google storage for the data is a benefit, although since the storage and notebook capabilities are not hosted in the same area, connecting to said data can be an issue. Colab provides access to common ML libraries which can be loaded into the environment, but uploading of user-generated libraries is limited. On the free version of Colab, GPU and TPU processing may be limited due to traffic on the site, and disconnections to the compute services can be frequent, especially if the user has stepped away for a short time. Data uploads to the Colab environment are not permanent, and connections disappear after the session ends, which can lead to confusion as to where the data are stored. Uploading large datasets to Google Drive can be slow and prone to dropout. The Colab platform stores files on Google Drive with 15GB of free capacity; however, dealing with larger datasets requires more space, which makes execution challenging. Google Colab sessions allow customers to connect to the service for up to 12 hours a day. To work for a longer period, customers can use the commercial version, Colab Pro, which allows programmers to remain online for 24 hours. Upgrading to the paid version of the service can ameliorate many of these concerns.

Commercial Options

Due to the complexity of the calculations needed, computer scientists and mathematicians have primarily led the advancement of ML and DL. The three major cloud-provision tech companies (Microsoft, Google, and AWS) have made the move to develop cloud-based ML tools that aid businesses to leverage AI. AWS is the most developed of the three, having been in the market for the longest time, however it is also the most expensive. Their pricing structure is complex, but extremely fine-grained, allowing the customer to purchase only the minimal amount of computing power needed to complete the task. Many of these commercial services offer resources as well as access to pre-trained models and Jupyter notebooks with prewritten code that can be easily modified to suit the required processing task. A free tier of services is available for one year, although users may create new accounts as needed to extend this access.

CONCLUSION

Currently, ML is heavily dependent on large amounts of data to be effective. For DL models to generalise and thus provide correct predictions for image classification, thousands of examples are needed. This can be a problem for researchers that do not have access to large data sets or data that are well labelled. In addition, ML algorithms can be complex to implement in code and often require a high level of programming knowledge. This can be a barrier for researchers who do not have the resources to invest in training. Furthermore, ML is often not able to provide accurate results in cases where the data is very noisy or unbalanced. This can be a problem in many real-world applications, such as medical diagnosis or fraud detection.

However, modern development tools are starting to flatten the learning curve necessary to do basic ML such as image classification. Although computer vision has existed since the 1960s, hardware requirements and computational knowledge have always made it highly complex and difficult to implement. This project suggests that archaeologists with limited resources have access to lower-code image analysis choices. Some platforms, such as Google and AWS, have begun to develop software for this purpose. The development of automated solutions for the identification of rock art and ceramics that extends beyond the simple binary classifications are now available. The experiments in this study have used labelled data and CNN to predict categories as well a pre-trained model with no labelling and using feature extraction, by using representations learnt by a prior network to extract relevant features from incoming data. In the instance of KCN, stylistic variance may indicate intersite and intrasite identification, as well as regional cultural links.

ACKNOWLEDGEMENTS

This work was made possible with the support of Dr Noel Tan and Katherine Cameron who provided access to their respective photo datasets of KCN and BNW. Archaeological work at BNW has been supported by the National Research Council of Thailand and the Thai Fine Arts Department and the kind cooperation of the BNW community. Additionally, the authors would like to express their gratitude to Asst. Prof. Patcharee Saributr and the Thai Fine Arts Department for their compiled website of 'Paintings, colours and engraved images Historical art in Thailand' (http://www.era.su.ac.th/RockPainting/index.html), without which this project would not have been possible. Parts of this paper were presented at the 22nd meeting of the Indo-Pacific Prehistory Association in Chiang Mai.

REFERENCES

Bickler, SH (2021) Machine learning arrives in archaeology. *Advances in Archaeological Practice* 9(2): 186-191.

Cameron, K (2013) Periphery to the centre: Social and Cultural Change over time at BNW, Northeast Thailand. Honours Thesis, James Cook University, Townsville

Davenport, TH and Mittal, N (2023) *All-in on AI: How Smart Companies Win Big with Artificial Intelligence*. Boston: Harvard Business Review Press.

El Naqa, I and Murphy, MJ (2015) What is ML? In: I El Naqa, R Li, MJ Murphy (eds.) *Machine Learning in Radiation Oncology*. Cham: Springer, 3-11.

Frye, M, Mohren, J and Schmitt, RH (2021) Benchmarking of Data Preprocessing Methods for Machine Learning-Applications in Production. *Procedia CIRP* 104: 50-55.

Harris, NJ and Tayles, N (2012) Burial containers–A hidden aspect of mortuary practices: Archaeothanatology at BNW, Thailand. *Journal of Anthropological Archaeology* 31(2): 227-239.

Higham, C and Higham, T (2009) A new chronological framework for prehistoric Southeast Asia, based on a Bayesian model from BNW. *Antiquity* 83(319): 125-144.

Higham, C (2011) The bronze age of Southeast Asia: New insight on social change from BNW. Cambridge *Archaeological Journal* 21(3): 365-389.

Higham, C (2014) *Early mainland Southeast Asia: from first humans to Angkor*, 1st edition. Bangkok: River Books.

Higham, CF and Kim, NC (2022) *The Oxford Handbook of Early Southeast Asia*. Oxford University Press.

Jalandoni, A, Zhang, Y and Zaidi, NA (2022) On the use of ML methods in rock art research with application to automatic painted rock art identification. *Journal of Archaeological Science* 144: 105629.

Jordan, MI and Mitchell, TM (2015) Machine learning: Trends, perspectives, and prospects. *Science* 349(6245): 255-260.

Keyes, CF (1967) Isan: Regionalism in Northeastern Thailand. Cornell Thailand Project; Interim Reports Series, 10. Ithaca: Department of Asian Studies, Cornell University.

Keyes, CF (1974) A note on the ancient towns and cities of northeastern Thailand. *Japanese Journal of Southeast Asian Studies* 11(4): 497-506.

Khanthakan, P (1979) Prehistoric Rock Art of KCN, Sikhio District, Nakhon Ratchasima Province. BA Thesis, Silpakorn University, Bangkok, Thailand.

Kowlessar, J, Keal, J, Wesley, D, Moffat, I, Lawrence, D, Weson, A, Nayinggul, A and Mimal Land Management Aboriginal Corporation (2021) Reconstructing rock art chronology with transfer learning: A case study from Arnhem Land, Australia. *Australian Archaeology* 87(2): 115-126.

Maaten, LVD, Boon, P, Lange, G, Paijmans, H and Postma, E (2007) Computer vision and ML for archaeology. Available at: https://lvdmaaten.github.io/publications/papers/CAA_2006.pdf [Accessed 7 November 2023].

Otárola-Castillo, E and Torquato, MG (2018) Bayesian statistics in archaeology. *Annual Review of Anthropology* 47(1): 435-453.

Pawlowicz, L, Downum, C and Terlep, M (2017) Applications of Machine Learning for Classification and Analysis of Southwestern US Decorated Ceramics, poster presented at the 82nd Annual Meeting of the Society for American Archaeology, Vancouver, British Colombia.

Prasomphan, S and Jung, JE (2017) Mobile application for archaeological site image content retrieval and automated generating image descriptions with neural network. *Mobile Networks and Applications* 22: 642-649.

Purshotam, A (2015) Automatic Indexing of South African Rock Art Images. Doctoral Dissertation, University of the Witwatersrand, Johannesburg.

Tan, NH (2014) Rock Art Research in Southeast Asia: A Synthesis. *Arts* 3: 73-104.

Tsigkas, G, Sfikas, G, Pasialis, A, Vlachopoulos, A and Nikou, C (2020) Markerless detection of ancient rock carvings in the wild: rock art in Vathy, Astypalaia. *Pattern Recognition Letters* 135: 337-345.

Widrow, B and Lehr, MA (1990) 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. *Proceedings of the IEEE* 78(9): 1415-1442.