# ARCHAEOLOGY IN NEW ZEALAND

# Machine Learning Identification and Classification of Historic Ceramics

**Simon H. Bickler**
**Bickler Consultants Ltd.**

## Introduction

This paper describes some preliminary results on using "machine learning" to assist in the identification of historic period ceramics recovered from archaeological sites in New Zealand. A large amount of time and energy goes into the identification of patterns found on the ceramics, almost all of it originally imported from the UK. Machine learning relates to the branch of computing which describes the study and programming of algorithms allowing computers to learn from data and then make predictions from that data(1) (see e.g., Shalev-Shwartz. and Ben-David). It underpins "artificial intelligence" and is being used in an ever-increasing number of fields leveraging the growth in computer power to solve a variety of problems. The algorithms are designed to allow computers to be "trained" to understand a set of data and then use that training to extract, classify, sort and draw conclusions of a new set of related data. Common uses of this approach include optical character recognition, where handwritten or scanned images of text are converted into digital readable text, as well as fingerprint identification, face recognition, and object identification in photographs and video. One well-known example, finding algorithms to accurately distinguish between photographs of cats versus dogs, is the basis of an on-going competition(2) to improve algorithms.

With historic ceramics, the goal is to determine whether the computer can identify a pattern, or provide a way of finding those patterns that are most similar. Image similarity is made up of a lot of factors relating to the structure of the image itself, for example, a whole object versus a sherd, the colours, the way in which the pattern is applied, and so on. Prehistoric pottery classifications commonly rely on formal classification schema which make matching easier (see Hörr *et al*., 2014). Although such schema are applied when discussing technique and function in the context of historic ceramics, using them for pattern identification is not undertaken because so many are actually named and mass produced. Basic keyword systems for instance can make significant difference to sorting the images, e.g., show patterns with flowers, birds, bands etc. This method relies on a time- consuming classification, which might involve the identification of hundreds of elements which may be present on a single pattern. This sort of pattern identification remains best done by people and their memory, but tools to improve the results and speed up the process can produce useful results for archaeological analysis.

Training computers to undertake the task of the pattern recognition and then classification on any sort of useful scale has been nearly impossible until recently. However, technological improvements in "computer vision" (see e.g., Bevan et al. 2014), teaching computers to interpret visual data, along with machine learning providing robust predictive trained models, allows computers to learn from large libraries of images. The final component has been making these abilities more widely accessible and that is the current revolution with major Information Technology companies investing heavily in providing these algorithms as services.

Four types of Machine Learning algorithms are demonstrated in this paper:

1. Template matching
2. Object identification
3. Deep Learning convoluted network – Image Similarity matrices
4. Trained Deep Neural networks pattern matching.

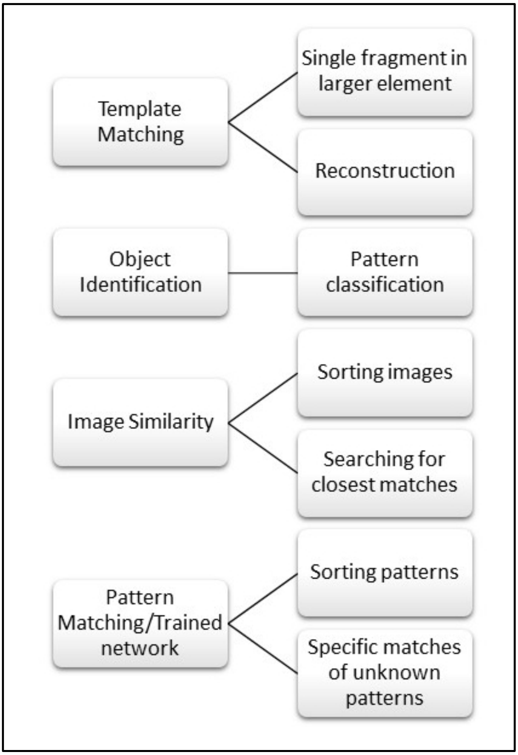


*Figure 1. Machine learning strategies and possible uses in archaeological ceramics.*

There is overlap in the techniques and the division here that relates to possible ways in which they can be used for analysing historic ceramic collections (Figure 1). The experiments were undertaken primarily on archaeological image collections held by the author and designed to show how they can be used in common tasks such as pattern identification and classification. The software used was derived from publicly available (and at least partly free) tools. Some examples required no programming while others involved using Python[3] code to access available machine learning "Accessible Programming Interfaces" (APIs). The advantage of using these APIs is that the "heavy lifting" involved with machine learning algorithms both in terms of processing power and sophistication did not require as much local computing knowledge or power. The downside is that none were specifically engineered for use with archaeological collections.

## Template Matching

Template matching algorithms are designed to search for a specific pattern within a larger set of data. The algorithm uses one image and then searches through another image or library of images for a match. The template can be as simple as a round or square object, a letter or number, a face or something more complex. The method is not as simple as it might sound as it must compensate for variations in scale, rotation and colour. Some may also be able to deal with images where part of the target may be obscured.

To demonstrate the algorithm, a small part showing the top of a tower in an image of a Rhine pattern sherd was used to create a "template" for matching (Figure 2). Open Computer Vision software[4] was used to search another Rhine image for examples of the tower. The resulting image shows where the software found matches in the target (Figure 2 right). Positive matches to the two similar towers in the target showed the effectiveness of the algorithm. However, the algorithm also identified several matches that were incorrect. Using different templates and fine tuning the search of the parameters does allow for better outcomes.

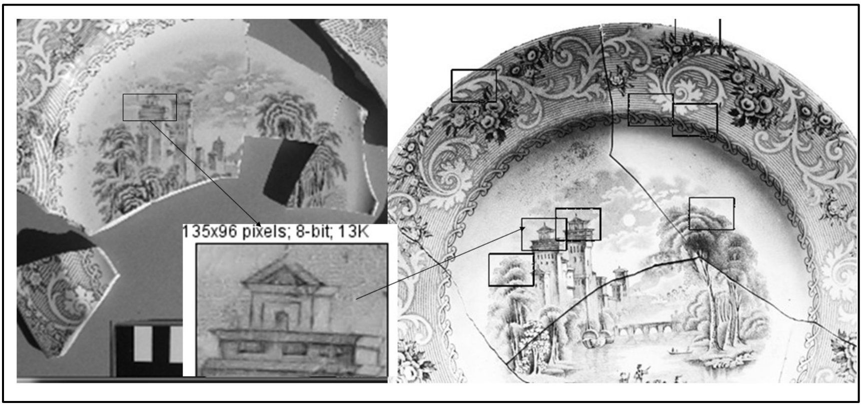


*Figure 2. Template matching from one Rhine sherd to another (left: source of template image: right results of template matching).*

Repeating the matching through a library of images, then scoring the number and certainty of the matches could then be used to find the most likely recorded pattern of the new fragment. There are difficulties because not only would each fragment have to be checked against thousands of patterns, but there can be considerable variation even within any specific pattern which may or may not be diagnostic. A pattern also can be made of multiple, and distinct, components while only part of a pattern may be present on a fragment.

An alternative use of the templating approach is to match fragments to a complete example (as in a jigsaw puzzle) to determine where a fragment belongs. Other applications in archaeology such as searching for archaeological features within a larger landscape using high resolution Digital Elevation Models (e.g., Jones and Bickler 2017) are being used.

## Object Identification

One of the most common uses for machine learning currently is the identification of "objects" in images. Object identification involves using a large library of images to "train" computers either to find different classes of objects within an image, e.g., a car, a person and so forth, or to distinguish between different "objects" e.g., cats versus dogs, which may be in image. The size of the training library is very large, usually in the order of thousands of images, so that the different elements such as colour, lighting, scale and perspective ensure that characteristics relating to a class of object can be identified and distinguished. Furthermore, the library images must be tagged by class so that the "training" is linked to each class of objects. Large image libraries available for training and testing of machine learning algorithms are regularly being created and can be used in a wide variety of applications. Unsurprisingly, none relate specifically to identification of historic ceramic patterns.

Using some of these general image libraries for machine classification of historic ceramics is instructive. A Rouen patterned plate was uploaded to Google Cloud Vision(5) (Figure 3 top). The algorithm did accurately determine that the image was probably that of a saucer or plate and "porcelain". The algorithm did not really focus on the pattern on the object. A second image showing just a transfer printed Near Eastern scene was uploaded to another service, Clarifai.co(6) (Figure 3 bottom). The highly ranked, i.e. most applicable, concepts it predicted within the scene included: "*print, people, illustration, art, woodcut, engraving, ancient, antique, old, painting, lithograph, watercraft, architecture, man, group, vintage*".

The recognition of a broad "style" of the pattern (e.g., "antique") means that with more work, the process could be effective in distinguishing between broad types of patterns. Object classification could also be used to identify features such as flowers, birds and animals, boats, etc. from the ceramic patterns in an automated manner. This would allow for some level of assistance in automated keyword creation and classification of images for databases.

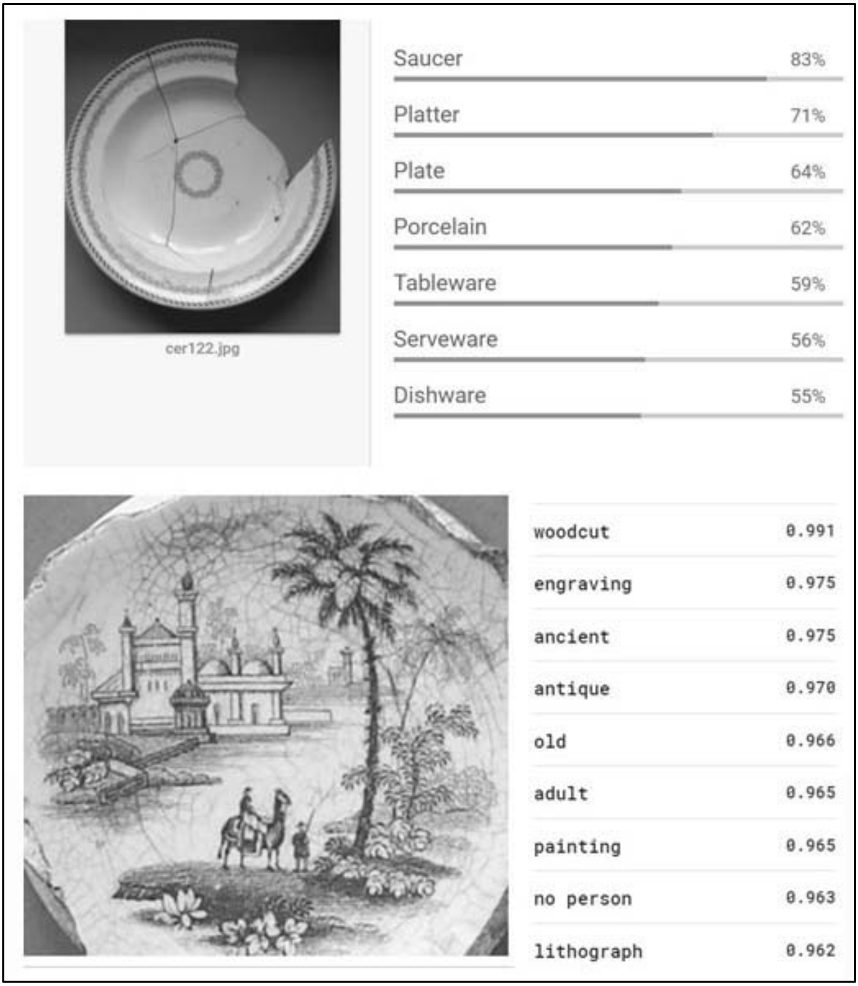


*Figure 3. Object identification on ceramic patterns; Top Google Cloud Vision interpretation of ceramic plate with Rouen pattern; Bottom: Scene from transfer print ware used in the Clarifai.com/demo classification.*

## Image Similarity

The previous results showed some promise but given the overall range and variation in the historic ceramic patterns, suggested that exact matching may be too ambitious. Another useful approach focuses on providing a methodology that takes an image and tries to find other *similar* ones. This approach relies on building up a library of patterns in images like those used in the object identification approach but rather than worrying about what specific objects are used in the patterns, the algorithms are just concerned with sorting out those images from the library according to how well they "look like" the pattern being examined. This general approach is like the template approach although the training algorithms create general descriptors of the library patterns and then compares those descriptors (vectors or tensors[7] in the parlance) to find a measure of "similarity". To illustrate this, a library of 100 images was collected with a range of transfer print wares. Initial experiments on the variety of images were not very successful in identifying patterns because the algorithms clustered images of whole plates, or mostly whole plates together regardless of the patterns because the algorithms placed emphasis on the shape of the object. To remove this effect, the testing database was created just using the borders. The border images were selected randomly to try to cover a range of different designs but included duplicates of some of the patterns.

The database of 100 border patterns was uploaded using the custom collection API from Indico.io[8] library using Python. Indico's custom image similarity algorithm uses pre-trained image libraries that are then modified with a library of the custom images for sorting. Crucially, this allows small collections of images to be analysed by the algorithm. This does sacrifice some potentially more accurate matching that a larger collection would provide but better reflects the challenge of archaeological samples which often have small numbers of ceramics with a diverse range of patterns. The experiment generates a similarity matrix for each image along with its most similar matches. A montage picture for each single image alongside its 9 most similar matched images identified was created including the measured similarity calculated by the custom classification. Two of the results are shown in Figure 4 and Figure 5. Figure 4 is based on a sherd with one of the more common patterns, Rouen, and the algorithm has closely matched it with the other Rouen sherds in the collection. Other patterns are similar and the similarity is likely due to the density of the border pattern rather than specific elements in it. Figure 5 shows matching with a sherd based on part of the Rhine pattern. Interestingly the closest match to that image was not Rhine at all, but another striped pattern. The matching algorithm emphasised the parallel line part of the Rhine border. The next closest match was another Rhine image and that matched despite rotation and scale differences between the images.

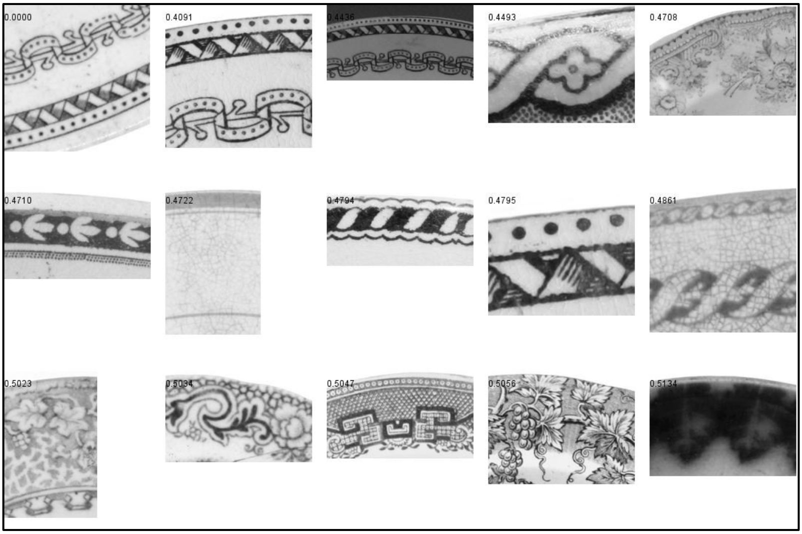


*Figure 1. Matching of Rouen pattern (top left source image) and nearest patterns.*

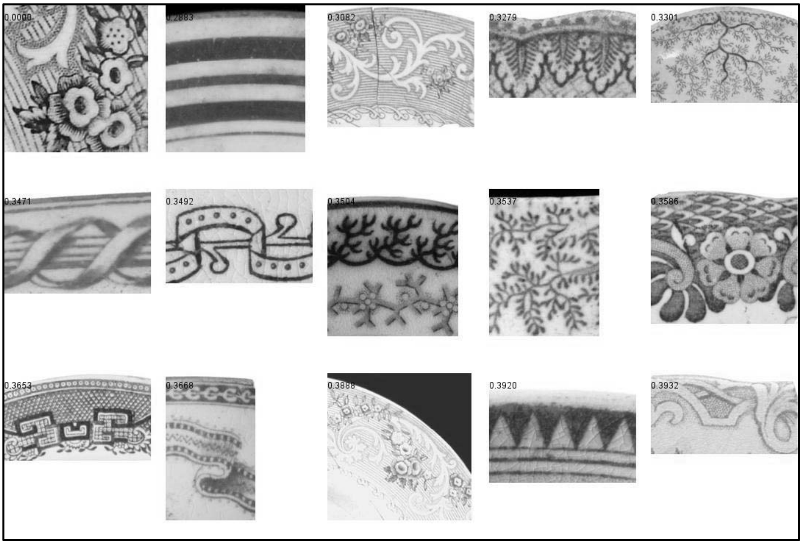


*Figure 2. Image similarity with closest matches to part of Rhine pattern motif.*

The algorithm is given no additional information about the patterns i.e., which ones are from the same pattern and the results are encouraging. The algorithm offered up a selection of pattern images that are most like the source image making visual confirmation straightforward. As the number of patterns and images increases, success will depend on 1) how good the coverage of the library is in capturing the range of patterns likely to be found, 2) finding different ways to process the border images, for instance, to provide more usefully diagnostic images (such as removing colour), and 3) testing out different similarity measures and thresholds to explore how to improve the outcomes.

## Pattern Matching

The holy grail of the machine learning approach is to get an algorithm to correctly identify the pattern on a sherd. This can be done by providing the machine learning algorithms a library of pre- classified images containing each pattern. As in the object identification approach, the algorithm looks at each set of images describing a named pattern and establishes a mathematical description of what it finds similar in each set. A new image or set of images is then compared with the pre- trained set of descriptions(9) and an estimate of how likely the pattern from the new image fits the pre-trained description is calculated.

An initial training library of 3 identified patterns was created and then expanded to 5 patterns for testing. Each pattern was characterised using only 10-15 images with a variety of shots of vessels and vessel fragments for each. This was designed to avoid the training focussing on the overall shape of the object in the image and emphasising the similarity of the shared pattern components. Some alternatives based on different colours were also included (and greyscale images were also tested). The collection of tagged images was then used to create a trained model using the same Indici.io API described above(10).

A second set of 22 images was then classified by the algorithm using the trained model to determine whether the algorithm could identify the correct pattern. The test images included a range of the patterns used in the model with a few additional patterns added to see how well the model would cope. The initial experiment with only 3 patterns was very successful with all of the images of the 3 patterns correctly assigned with a high likelihood. Images of patterns not in the original 3 were all assigned to a test pattern with varying, but low, probability. The process was repeated with the 5- pattern model and the results are shown in Table 1. This shows the actual designated pattern of each test image and the resulting probability match to each of the assigned pattern.

*Table 1. Results of testing images compared to a 5-pattern trained library (Shaded cells indicates correctly identified results where the highest probability of match is indicated in bold in each row).*

| Image | Actual | Asiatic Pheasants | Cable | Rhine | Rouen | Willow |
|---|---|---|---|---|---|---|
| britpots-1044 | Cable | 0.030 | **0.574** | 0.082 | 0.019 | 0.296 |
| britpots-1048 | Willow | 0.011 | 0.056 | 0.148 | 0.008 | **0.776** |
| britpots-1064 | Asiatic | 0.293 | 0.055 | **0.417** | 0.024 | 0.211 |
| cer91 | Rhine | **0.738** | 0.060 | 0.072 | 0.043 | 0.088 |
| fhwh12 | Willow | 0.069 | 0.038 | 0.160 | 0.052 | **0.680** |
| kaw_0394 | Willow | 0.037 | 0.040 | 0.003 | 0.010 | **0.910** |
| lotus | Lotus | 0.008 | 0.018 | 0.419 | **0.543** | 0.012 |
| Oashore_0001 | Asiatic | **0.711** | 0.003 | 0.266 | 0.009 | 0.012 |
| Scott_0043 | Cable | 0.010 | 0.144 | 0.041 | **0.789** | 0.016 |
| Scott_0044 | Cable | 0.021 | 0.053 | 0.073 | **0.848** | 0.005 |
| Scott_0045 | Cable | 0.111 | 0.101 | 0.065 | **0.714** | 0.009 |
| Scott_0200 | Willow | 0.009 | 0.005 | 0.212 | 0.018 | **0.756** |
| Settlers_0009 | Rhine | 0.269 | 0.032 | **0.634** | 0.015 | 0.051 |
| Settlers_0026 | Willow | 0.024 | 0.004 | 0.170 | 0.024 | **0.779** |
| Settlers_0027 | Willow | 0.005 | 0.004 | 0.066 | 0.013 | **0.911** |
| Settlers_0028 | Willow | 0.071 | 0.169 | 0.104 | 0.047 | **0.609** |
| TeHoe_0016 | Cable | 0.055 | **0.336** | 0.157 | **0.413** | 0.040 |
| VPT_0014 | Rouen | **0.506** | 0.026 | 0.108 | 0.355 | 0.006 |
| WellingtonBypass_0565 | Rhine | **0.814** | 0.019 | 0.056 | 0.034 | 0.077 |
| WellingtonBypass_0567 | Rhine | 0.164 | 0.016 | **0.701** | 0.038 | 0.080 |
| Westney_0047 | Asiatic | 0.223 | 0.012 | **0.345** | 0.108 | **0.312** |
| wyn_0003 | Asiatic | **0.524** | 0.012 | 0.215 | 0.012 | 0.237 |

*Table 2. Sample of results of 100 border patterns classified to a 5-pattern trained library (Shaded cells indicates highest value).*

| Image | Actual Pattern | Asia. Pheasants | Cable | Rhine | Rouen | Willow |
|---|---|---|---|---|---|---|
| Abbeville_0008 | A. Pheasants | 0.899 | 0.061 | 0.024 | 0.006 | 0.010 |
| britpots-1055 | A. Pheasants | 0.850 | 0.041 | 0.076 | 0.010 | 0.023 |
| Oashore_0002 | A. Pheasants | 0.734 | 0.008 | 0.187 | 0.009 | 0.062 |
| Westney_0001 | A. Pheasants | 0.961 | 0.006 | 0.025 | 0.003 | 0.005 |
| Abbeville_0005 | Banded | 0.617 | 0.219 | 0.004 | 0.120 | 0.040 |
| Blomfield_0074 | Cable | 0.021 | 0.892 | 0.011 | 0.009 | 0.068 |
| britpots-1043 | Cable | 0.015 | 0.796 | 0.011 | 0.006 | 0.171 |
| ButlerStoney_0001 | Cable | 0.567 | 0.294 | 0.002 | 0.010 | 0.127 |
| Settlers_0032 | Cable | 0.764 | 0.145 | 0.004 | 0.027 | 0.059 |
| TeHoe_0025 | Cable | 0.229 | 0.631 | 0.031 | 0.076 | 0.033 |
| UCOL_0088 | Cable | 0.131 | 0.434 | 0.012 | 0.347 | 0.076 |
| UCOL_0088a | Cable | 0.824 | 0.139 | 0.020 | 0.009 | 0.008 |
| UCOL_0089 | Cable | 0.071 | 0.270 | 0.021 | 0.237 | 0.401 |
| UCOL_0089a | Cable | 0.276 | 0.637 | 0.065 | 0.003 | 0.018 |
| Blomfield_0081 | Coral | 0.137 | 0.288 | 0.141 | 0.005 | 0.429 |
| Blomfield_0081a | Coral | 0.638 | 0.238 | 0.034 | 0.009 | 0.081 |
| Athenree_0001 | Fibre | 0.298 | 0.219 | 0.195 | 0.064 | 0.224 |
| Blomfield_0034 | Fibre | 0.829 | 0.047 | 0.028 | 0.029 | 0.066 |
| Scott_0011 | Fibre | 0.723 | 0.037 | 0.028 | 0.002 | 0.211 |
| Westney_0018 | Fibre | 0.308 | 0.151 | 0.265 | 0.011 | 0.265 |
| Blomfield_0095 | Marine | 0.083 | 0.774 | 0.127 | 0.013 | 0.002 |
| Blomfield_0095a | Marine | 0.786 | 0.157 | 0.051 | 0.002 | 0.005 |
| Blomfield_0107 | Rhine | 0.038 | 0.053 | 0.776 | 0.074 | 0.058 |
| cer79 | Rhine | 0.060 | 0.017 | 0.882 | 0.007 | 0.034 |
| Scott_0019 | Rhine | 0.050 | 0.151 | 0.724 | 0.029 | 0.045 |
| TeHoe_0008 | Rhine | 0.081 | 0.002 | 0.813 | 0.031 | 0.073 |
| Westney_0044 | Rhine | 0.037 | 0.013 | 0.888 | 0.011 | 0.051 |
| Blomfield_0109 | Rouen | 0.020 | 0.253 | 0.087 | 0.332 | 0.308 |
| Blomfield_0109a | Rouen | 0.217 | 0.583 | 0.022 | 0.004 | 0.175 |
| cer140 | Rouen | 0.006 | 0.084 | 0.015 | 0.264 | 0.632 |
| Scott_0023 | Rouen | 0.028 | 0.055 | 0.010 | 0.175 | 0.732 |
| jud1 | Willow | 0.353 | 0.065 | 0.043 | 0.003 | 0.535 |
| Rangiriri_0010 | Willow | 0.723 | 0.028 | 0.016 | 0.025 | 0.209 |
| Rangiriri_0011 | Willow | 0.428 | 0.016 | 0.013 | 0.157 | 0.386 |
| Well. Bypass_0128 | Willow | 0.010 | 0.005 | 0.007 | 0.005 | 0.973 |
| Well. Bypass_0532 | Willow | 0.014 | 0.006 | 0.009 | 0.006 | 0.964 |

While most examples belonging to the trained patterns were correctly identified, and the probability of the match was high, this was not always definitive. Most of the sherds with patterns not included in the model matched poorly (probability < 0.5) but some examples gave better matches than those from patterns that were in the training database. Even a minor shift from 3 to 5 patterns immediately reduced the effectiveness of the matching. This is problematic given that the goal is to be able to eventually work on thousands of different patterns, but demonstrated the need for large libraries of training images that capture the variation of each pattern and the range of different patterns.

The experiment was re-run with the 100 border images used in the image similarity experiment. A sample of the results is shown in Table 2. The results matched the Asiatic Pheasant and Rhine patterns consistently, were mixed for Cable and Willow but very poor for Rouen. Most Rouen images were classified as Cable. This reflected a broad visual likeness of the patterns (linear band), and could be improved by increasing the quality and number of Cable and Rouen examples in the testing.

Overall, the results showed that the machine learning could identify patterns with the right images from which to train. Reliable classifiers often use hundreds and thousands of images, but for archaeologists this is a major limitation because the patterns found frequently are easily identifiable and there is little value in having the machine learning algorithms work hard to find pattern matches that take no significant effort for people to identify. One objective for future work is expanding the image libraries for the patterns and ensure that images of rare patterns are better described by re- using available images in multiple ways, e.g., rotating, cropping, and so forth to allow the training to distinguish what makes it unique.

## Discussion

The results presented here are preliminary but demonstrate the potential for machine learning algorithms to significantly assist in the study of archaeological artefacts. Four different approaches summarised in Figure 1 have been described to show some of the possibilities in using the algorithms with classification tasks and particularly in assisting in the identification of transfer-print wares from New Zealand historical archaeological sites. Utilising easily accessible libraries and algorithms shows what can be accomplished and provides a guide as to where future work should be aimed.

The machine learning algorithms can assist in the often-times mammoth task of sorting, managing and analysing large historic ceramic collections from archaeological sites. Currently the image similarity approach seems to provide the best results but the pattern matching has potential. To make this more effective, the current databases from sites would need to be combined and used to provide the basis of the “training”. It will be necessary to use some of the cloud-based services to bring the necessary computing power to bear on the problem, but the cost and availability of these services has dropped to the point where this is no longer the barrier that it once was.

The available algorithms for pattern identification are becoming increasingly accessible and powerful and the task is to bring these to bear more specifically on identifying and classifying patterns in an archaeologically useful manner. The best results are likely to involve using standardised classification methods such as keyword descriptions, manufacturer information, techniques and colours, combined with machine learnt similarity or pattern matching algorithms. A service which captures an image of a newly found ceramic artefact and uploads it to an online service for preliminary identification and classification either via mobile device or desktop pc is feasible in the near future.

Machine learning is not the beginning of the end for human specialists who still have the upper hand in archaeological tasks. As historian Harari (2017) recently suggested “[t]here are some safe jobs: the likelihood that algorithms will displace archaeologists is only 0.7 percent”. We are unlikely to be replaced soon but we may well be able to get some much-needed assistance from some well-trained artificial intelligence.

## Acknowledgements

I would like to thank Dorothy Brown, Ben Davies and Jennifer Low for their editorial assistance. The images used in the report come from several projects and I would like to thank Jaden Harris, Mica Plowman and Jennifer Low for the images and their work on the historical ceramics database over the years.

## Endnotes

1. https://en.wikipedia.org/wiki/Machine_learning
2. https://www.kaggle.com/c/dogs-vs-cats
3. Python is just one computer language of a wide range of programming languages available for machine learning applications
4. Programmed in Python using Open CV version 3 (http://docs.opencv.org/3.1.0/d4/dc6/tutorial_py_template_matching.html)
5. https://cloud.google.com/vision/ Accessed May 2017
6. https://www.clarifai.com/demo Access May 2017

7. These are mathematical descriptions of the information from the analysis of the images that describe the classification
8. https://indico.io Custom Collection API Accessed March 2017
9. Referred to as a vector or tensor, which is a mathematical formula of the image or pattern
10. The Microsoft CustomVision API https://www.customvision.ai/ provides a similar functionality

## References


Bevan, A., X.J. Li, M. Martinón-Torres, S. Green, Y. Xia, K. Zhao, Zh. Zhao, Sh. Ma, W. Cao and T. Rehren. (2014) Computer vision, archaeological classification and China's Terracotta Warriors. *Journal of Archaeological Science* 49: 249–54.

Harari, Y.N. (February 2017) The rise of the useless class. http://ideas.ted.com/the-rise-of-the- useless-class accessed March 2017.

Hörr, C., Lindinger, E., and Brunnett, G. (2014). Machine learning based typology development in archaeology. *Journal on Computing and Cultural Heritage* 7 (1): 1-23.

Jones, B. and Bickler, S. H. (2017) High resolution LiDAR data for landscape archaeology in New Zealand. *Archaeology in New Zealand*, 60(3): 35-44.

Shalev-Shwartz, S. and S. Ben-David (2014) Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press. Available online http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning- theory-algorithms.pdf.