ARCHAEOLOGY IN NEW ZEALAND

Prospects for Machine Learning for Shell Midden Analysis

Simon H. Bickler
Bickler Consultants Ltd

Introduction

This paper describes using machine learning (ML) techniques to assist in analysis of shellfish midden remains from archaeological sites in New Zealand. Midden analysis is the most common post-excavation analysis undertaken in prehistoric archaeological projects. Machine learning (ML) tools present the possibility of moving some of that work to automated systems which would then allow more time to be focused on the most complex and specialised tasks associated with such investigations. This work builds on discussion of such techniques used for the identification of historic ceramic patterns (e.g., Bickler 2018, this volume) and site identification from elevation data (e.g., Jones and Bickler 2017). No detailed discussion is presented here on the range of possibilities and importance of midden analysis as these are well covered in numerous publications in the New Zealand context (e.g., HNZPT 2014 and references listed there). The objective is to demonstrate how ML can assist with midden analysis although any useful tool will require additional hardware and software developments to get close to the level of expertise required for specialist work. Shell identification is not particularly difficult, but counting large quantities of shell is time-consuming.

Preliminary experiments in using ML for midden analysis shown here involve two separate types of analysis. The first is the creation of relevant images of shell samples for identification, with the goal of allowing scaling for analysing excavation midden in the future. The second uses ML to undertake the processing of the images to produce useful data for typical midden analysis such as species identification, counts and size estimates. The process described here involved:

1. Training a computer to identify different species of shell commonly found in New Zealand archaeological sites using collections of images;
2. Testing the species identification model;
3. Using an image of the shells to be identified;
4. Segmentation of individual shells from the combine sample image;
5. Species identification of each shell image;
6. Count of each shell species;
7. Evaluation.

The analysis was undertaken in the Microsoft RClient (part of the Microsoft Machine Learning Server or MMLS) environment to the R Statistics package (cran.r-project.org) and the Microsoft Machine Learning Server. The software and examples used are available free and this makes it easier and more cost-effective for future development by archaeologists.

Sorting and Segmentation

The first part of the analysis involves using tools to segment the images of individual shells from photos of a collection of shells, such as a midden sample. This process does not involve any ML but the algorithms used to individuate objects from images derive from general computer vision (CV) analysis with many different solutions. The method adopted here derives from software used to identify cells in microscopic images using a R statistics package called EBImage (Pau et al. 2010; available from bioconductor.org).

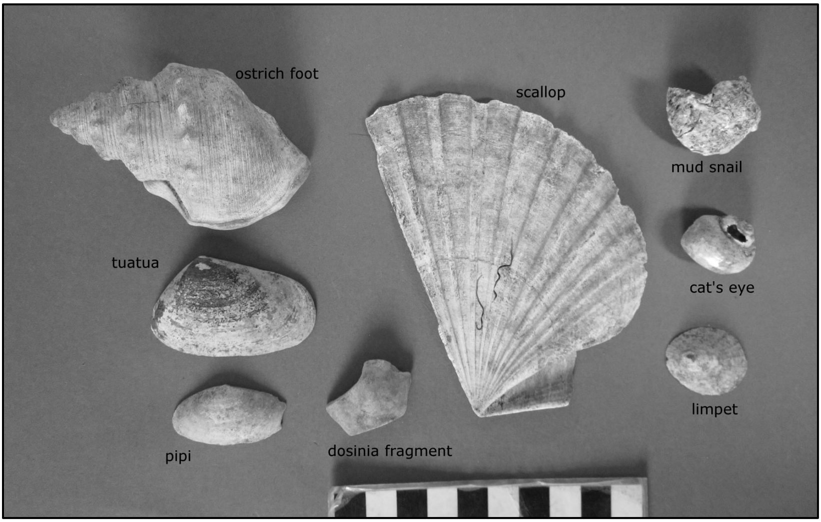

Figure 1. Photograph of "midden sample" showing a range of species used to test the ML model

An example image (Figure 1) showing a range of species found in the midden sample was analysed to distinguish each shell separately. The masking algorithm looks at areas of bright colour (i.e. white) versus a dark background colour and creates a "mask" that bounds each high contrast area and identifies that as a separate object, in this case a shell. Fine tuning the masking image analysis is specific to the lighting and exposures of each photo, but that relates as much to a lack of standardisation of setup and could be remedied. Once the masking of the individual shells is accurate, each separate object/shell image is saved as a new image for later identification (Figure 2).

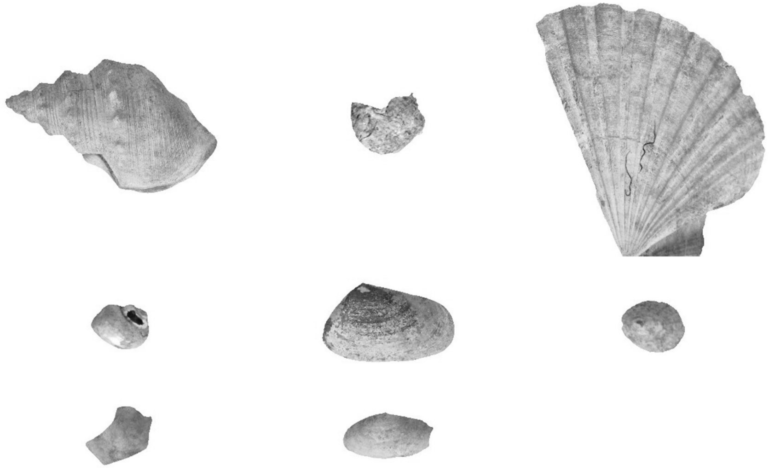

Figure 2. Segmented images of shells from the "midden sample."

Several issues are immediately apparent. Each shell must be separated out in the group photograph so that it can be masked properly. It is possible for overlapping shells to be separated but the process of segmentation becomes more difficult and unreliable. There are solutions for getting around the problem, such as large sorting trays, and conveyor belts which could be used to ensure each object is identified quickly and efficiently, but this is beyond the scope of the current work.

A more difficult problem relates to managing fragmentary shell. Obviously, the more fragmented the shell, the more difficult and unreliable any later automated identification and analysis will be. However, there are elements of that issue that can be explored using the CV tools. Figure 3 shows a photograph of a small pile of cockle shell, distinguishable but close together. The masking was designed to centre on each object but not necessarily go fully out of the boundaries of the object, as this could mean that neighbouring objects would be grouped together. Each segmented shell was identified and the automated count of "objects" matched the number of shells. Other approaches such as using multiple images of a pile of shells, for instance, could give a volumetric value and along with 2D image counts and size estimates. An automated system could then provide a reasonable estimate of the number of shells, level of fragmentation, potentially, and hence MNI which could be compared with weight and other methods of bulk sample analysis.

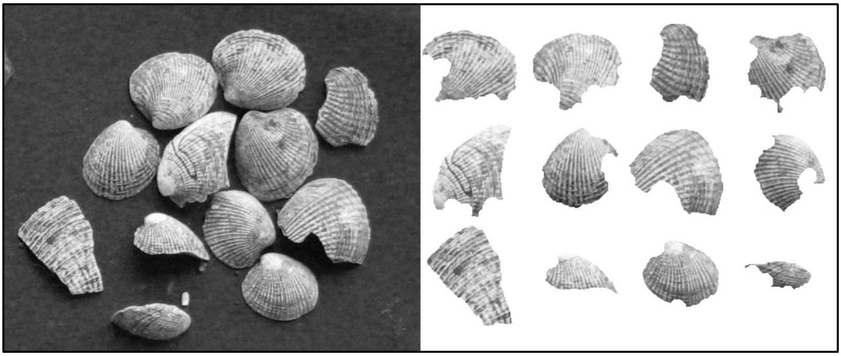

Figure 3. Cockle pile from midden (left) and segmented images of identified individual cockle shells.

One bonus of the image segmentation is that it is possible for the size of the shell to be calculated. It is straightforward to calculate the size of each shell identified by measuring the area of the shell image created by the mask. While this is in "pixels", as long as a photo-scale is included in the larger image (and no other scaling occurs), it is easy to convert that measurement into square mm or cm. As well as area, the maximum length and width can also be calculated and recorded. No testing has been done to explore the accuracy of these measurements compared with manual measurement, but for some species the "area" measure on a 2D plane may be a better size indicator than just length, which is the commonly used approach.

Machine Learning

The next stage is building a trained ML model used to identify shell species. ML refers 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 (see e.g., Shalev-Shwartz and Ben-David 2014). Broadly speaking, ML uses statistical techniques to analyse a set of categorised "training" data and derive a series of mathematical descriptions ("descriptors" or "feature vectors") for each of the images. Ideally, each category of object is therefore mathematically distinct from all other categories so creating a trained model is all about finding the best set of mathematical "features" for accurately identifying each example for all the categories. Archaeologists familiar with statistical techniques such as principal components analysis and discriminant analysis (see for example in obsidian sourcing studies such as McCoy and Carpenter 2014), use many of the same algorithms to create the descriptors or feature vectors. In the case of image data, as used here, each image is converted into a set of feature vectors, "image featurisation", and these are then analysed to create the groups of different images, in this case different shell species.

Creating good models typically requires thousands of images, but for most archaeological purposes that can be difficult to achieve. The approach taken here uses a pre-trained deep neural network (DNN) model to extract relevant features from images, so the model has some information about how to cope with image data (relating to distinguishing different objects in images) and then adding the smaller library of images relevant to the specific task, shells, to build image features that are specifically relevant (see Horton and Paunic 2017, Shalev-Shwartz and Ben-David 2014: 268ff).

Species Identification Model – Image Featurisation

The training of the shell identification model was done using the image featurisation abilities of the Microsoft Machine Learning Server software. The MMLS allows for use of a pre-learned DNN model, in this case the ResNet 101 model (He et al. 2015). This is one of the larger models available and based on thousands of small images, 224 pixels in size that has already been analysed to create a set of image features. The MMLS server is accessed via the R Statistics package, which then also manages the featurisation of the shell library and test sample images and combines the results with the segmentation information.

The methodology here follows in part from an example classifying the morphology of different wood knots (Horton and Paunic 2017) which relates to grading timber. This shows how flexible the ML approach can be when applied. The machine learning models are rapidly evolving, and new options will no doubt change and improve on the reported example.

Ten species were chosen for this experiment to create a trained model for shell identification. The species included those commonly targeted by Maori during prehistory including, pipi (*Paphies australis*), cockle (*Austrovenus stutchburyi*), ostrich foot (*Struthiolaria papulosa*), tuatua (*Paphies subtriangulata*), mudsnail (*Amphibola crenata*), mud whelk (*Cominella glandiformis*), limpet (*Cellana radians*), cat's eye (*Turbo smaragdus*), ringed dosinia (*Dosinia anus)* and scallop (*Pectin novaezelandiae*). The selection allowed for some likely confusion with shells such as the pipi and tuatua and the cockle and dosinia which are similar in shape but not size, as well as a mixture of univalves and bivalves.

A library of training images was then created for each of the 10 species. The target was to create around 20 images for each species using archaeological reports and the web although that number varied between 8 for limpets and 27 for ringed dosinia and tuatua. A total of 191 images were used for the initial model. The variability was purposeful because testing the robustness and flexibility of the model was part of the investigation.

The training images were processed to remove any background information. Some images were created from base images by rotation and mirroring the original image to bring the numbers up to ensure that the model was trained on the shape from different angles to make sure orientation was not likely to matter in later species identification.

The image featurisation was then carried out using the MMLS server in R (rxFeaturize function). This includes converting the images to match those of the Resnet 101 image library in size and then extracting the feature vectors that represent each of the shell images. This image featurisation library can be saved and reloaded, as well as expanded or altered to include new images and species.
The next step was to examine how well the model works with respect to distinguishing between the different species. The featurised data is randomly split (typically around 2/3 to ¾) to go into the training set with the rest into a test sample. A "Random Forest" algorithm is then used to fit the feature vectors to the species identification for the training samples (using R package randomForest package [Liaw and Weiner 2002]), which means that that set describes what each species "looks like". The test data is then tested against the "fitted data" and the probability of each sample belonging to a species is calculated. The accuracy of the tested model can then be examined to determine how well the classification worked. An example of the results can be summarised in a "confusion matrix" (Figure 4) which compares the known species against the predicted species of the images from the test sample. Around 85% of the test images were accurately identified, but the result varies between around 75-90% as the choice of training data and testing data is changed.

The result is not perfect but given the small library of training data, very promising. Each column shows the number of shell images from the testing group comparing which species they were predicted to be in versus the one to which they actually belong. Ideally, each species column will be single and in their correct class, which most of them are except for cat's eye, ringed dosinia and tuatua (in the example shown in Figure 4). As expected, depending on the sample selection some images of pipi or tuatua are confused but overall the model gets most of them correct. Other confusions such as cockle and the ringed dosinia are similarly mixed up at times because when the images are scaled to a uniform size they are indeed quite similar. This may also explain confusions between cat's eye, mud whelks and mud snails. Some confusion also lies in the difference between images of shells in good condition and worn examples found in excavations.

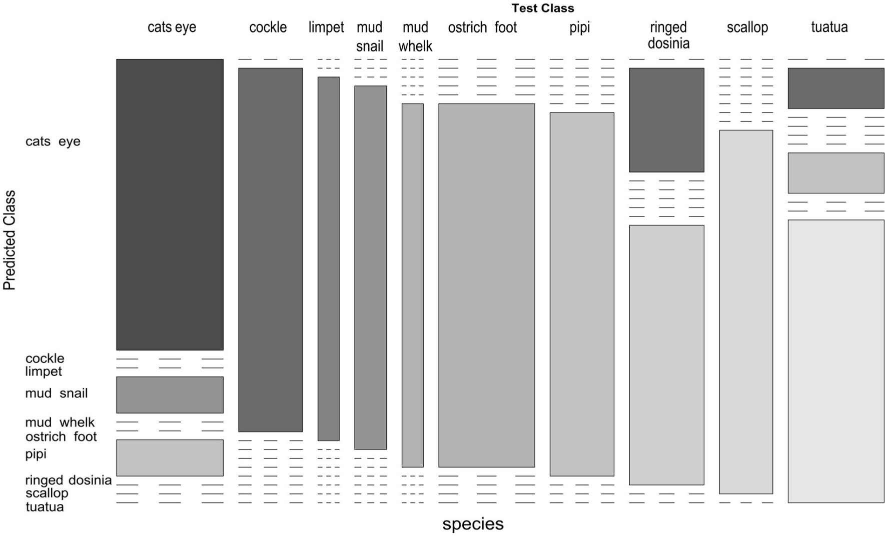

Figure 4. Mosaic plot showing the confusion matrix of a test sample of a 1/3 of the data.

Putting It All Together

Despite the deficiencies in the trained model, the results were considered good enough for analysing the midden sample. The individual shell images from the midden sample (Figure 2) were then "featurised" using the same process of the library data described above to create the species identification model. Furthermore, the masked area of each shell was measured, along with other parameters, scaled based on the original image scalebar, and the area of each shell calculated and saved for the later analysis. The random forest classification was then applied once more and a predicted probability for each shell match to a species was calculated. The results are shown in Figure 5 and summarised in Table 1. Despite the complexity, these steps do not take a standard computer more than a few minutes to complete.

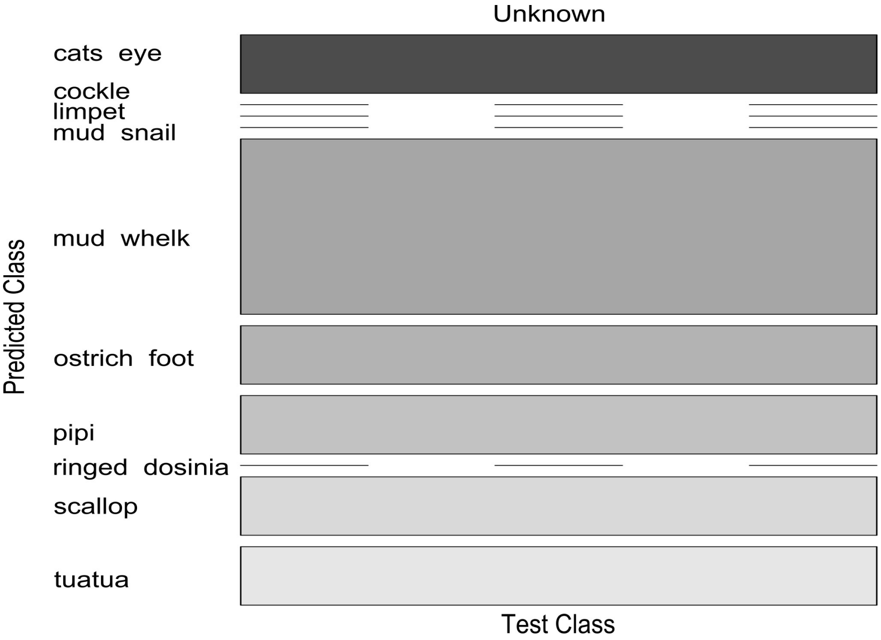

Figure 5. Mosaic plot showing categorisation of "unknown" midden sample species identification

The results are good given the quality and partial fragmentation of the samples. The ostrich foot is correctly identified and pipi, tuatua distinguished, and despite the scallop shell being damaged, it is also correctly identified. The other species are less reliable but given some damage, fragmentation and poor quality of the images of smaller shells, the results are still promising. The probabilities shown in Table 1 reflect the range of reliability of the predicted classification. Those values can be further analysed to provide the process of secondary analysis that might focus on distinguishing more difficult samples using more focused models.

Table 1. Result of species identification of midden sample using the ML model showing predicted probabilities of each shell to species and estimated shell area. Highlighted values show best match.

Known species	Cats eye	Cockle	Limpet	Mud snail	Mud whelk	Ostrich foot	Pipi	Ringed dosinia	Scallop	Tuatua	Predicted species	Shell area (cm^2)
Ostrich foot	0.102	0.084	0.04	0.108	0.152	**0.308**	0.042	0.058	0.062	0.044	Ostrich foot	25
Mud snail	0.152	0.086	0.064	0.106	**0.186**	0.112	0.078	0.052	0.082	0.082	Mud whelk	5
Scallop	0.052	0.082	0.046	0.054	0.066	0.14	0.05	0.072	**0.366**	0.072	Scallop	62
Cat's eye	**0.152**	0.066	0.062	0.12	0.144	0.05	**0.152**	0.068	0.068	0.118	Cats eye/pipi	3
Tuatua	0.12	0.064	0.05	0.114	0.13	0.06	0.138	0.032	0.08	**0.212**	Tuatua	12
Limpet	0.118	0.118	0.146	0.098	**0.15**	0.06	0.084	0.07	0.07	0.086	Mud whelk	4
Dosinia	0.096	0.056	0.042	0.13	**0.208**	0.178	0.062	0.078	0.076	0.074	Mud whelk	4
Pipi	0.09	0.116	0.11	0.104	0.108	0.038	**0.142**	0.106	0.058	0.128	Pipi	5

Discussion

This paper demonstrates a first attempt at creating a tool for automated midden analysis using ML techniques and shows these techniques can be used in archaeological analysis of midden. The current process requires using photographs or video to record the midden which is a hurdle, although some additional benefits might accrue in creating a detailed digital record of the midden and to get additional morphological data from the shells.

Another issue is that there are many species of the shellfish that are only minor variants of each other that are detectable by an expert but not necessarily easy to train from images. Where these species are from geographically distinct areas, the archaeologists could reclassify their results appropriately without much problem, but otherwise more detailed model building would be required. Creating region specific models that focus on the species likely to be found in a sample rather than using a one-size fits all model is one possibility. Misclassification of species can also be minimised by pre-training models of the species of shells identified from a sample of the actual midden and eliminating those shells species that are not present. This means that limiting the choices available will make the model less likely to produce false results with an emphasis then on the counting and sizes of those species known.

Another problem that arose from using the pre-trained models for the shell was that it eliminated size as a significant morphological trait. This was particularly relevant during the model building which standardised all the images for analysis to around 220+ pixels on each side. The effect of this is to remove shell size as a variable which meant species that are morphologically similar but different in size were confused. Juvenile specimens would also then become a factor. A larger and improved library of species images should greatly improve the model.

The experiments shown here demonstrate that ML represents a viable process for improving midden analysis. The motivation for improving the speed of midden analysis that techniques such as ML tools offer, relate to the demands of ensuring that sufficiently large samples of midden are analysed quickly and efficiently. The example demonstrated here is one process and new tools are rapidly changing the ease and effectiveness of ML application development. The analysis of other faunal remains, tool morphology and non-image based data such as chemical sourcing concentrations are just some of the other applications where ML could be effective. ML techniques are likely to transform analysis for prehistoric archaeology in the future.

Acknowledgements

I would like to thank Dorothy Brown and Jennifer Low for the editing and comments.

References

Bickler, S. H., (2018) Machine learning identification and classification of historic ceramics. *Archaeology in New Zealand*, 61 (1): 21-33.

He, K., X. Zhang, S. Ren and J. Sun, (2015) Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385.

HNZPT, (2014) Guidelines for Midden Sampling and Analysis, Heritage New Zealand Pouhere Taonga Archaeological Guidelines Series. Available at www.historic.org.nz.

Horton, R. and V. Paunic. (2017) Featurizing Images: The Shallow End of Deep Learning. Available at http://blog.revolutionanalytics.com/2017/09/wood-knots.html.

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

Liaw, A. and M. Wiener, (2002) Classification and regression by randomForest. *R News* 2 (3), 18-22.

McCoy, M.D. and J. Carpenter, (2014) Strategies for obtaining obsidian in pre-European contact era New Zealand. *PLoS ONE* 9 (1): e84302. doi:10.1371/journal.pone.0084302

Pau, G., F. Fuchs, O. Sklyar, M. Boutros and W. Huber, (2010) EBImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics,* 26 (7): 979–81.

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.