USING SUPERVISED MACHINE LEARNING TO CLASSIFY CERAMIC FABRICS.

by

Oksana Koval

A thesis submitted to the Faculty of Graduate Studies of

the University of Manitoba

in partial fulfilment of the requirements of the degree of

Master of Arts

Department of Anthropology

University of Manitoba

Winnipeg

Abstract

Objective: The purpose of this thesis was to assess the ability of supervised machine learning to discriminate between images of ceramic fabrics and evaluate the requirements for creating an effective dataset.

Method: Weighted Neighbour Distance using Compound Hierarchy of Algorithms Representing Morphology (*wndchrm*) algorithm, was applied to Zulu ceramic fabrics from South Africa. *Wndchrm* was used to extract thousands of image content descriptors, assign weights to the extracted features by learning their discriminative power from training examples, and classify unlabelled images by searching for a class with the nearest distance to the mean of the feature vector.

Results: Wndchrm was successful in distinguishing ceramic fabrics by differences in paste. Comparable results were obtained in separate experiments designed to identify differences in paste by region, community and individual potters. In all cases, a sample size of 50 training images per class was sufficient to produce 90% to 95% accuracy. In contrast, the experiments meant to identify fabrics by shaping techniques, represented by coiling and slab building, reached the accuracy of only 65% to 70%.

Conclusion: The experiments show that *wndchrm* is an effective method for classification of ceramic fabrics, which can increase consistency in comparing and classifying ceramic fabrics by variation in paste. Creating and sharing training dataset libraries is the next step necessary for wide adoption of supervised machine learning in this field.

Acknowledgements

I would like to express my sincere gratitude to Dr. Kent Fowler for his ongoing support and encouragement, for his indispensable depth of knowledge, and for believing in this project. I would also like to thank Dr. Mark Lawall and Dr. Robert Hoppa for their valuable feedback and interst in this work.

I dedicate this thesis to Henry.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis evaluates the effectiveness of the pattern recognition algorithm called *wndchrm*, which stands for the Weighted Neighbour Distance using Compound Hierarchy of algorithms Representing Morphology, to classify images of ceramic fabrics based on the variability in clay processing, paste preparation, and shaping techniques. Ceramics obtained from modern day potters have been used to cross-validate the results. The best practices for building an effective dataset to analyse ceramic fabrics using a supervised machine learning approach were also investigated. The findings presented in this thesis provide guidelines for dataset design and practical recommendations for future studies in this area.

1.1 Thesis Motivation

Cognitively grounded by our past experiences and ideas humans are limited in our ability to look at data objectively. Additionally, we are also physically constrained in how much information we can process at once in a meaningful way. Machine learning is an emerging field that is quickly gaining a cross-disciplinary traction as it can a) automate analysis and expedite research and progress; b) provide insights into large volumes of multi-dimensional data that is beyond

humans natural capacity to process; and c) widen the complexity of questions we can ask and problems that we can solve. Machine learning can also help increase the objectivity of data analysis. Before any data can be interpreted, it must first be ordered by units of information that can be easily compared and classified into discrete groups using specific criteria or clustered by similarity without assigning it into labelled groups (Banning, 2000). An archaeologist can make an unlimited number of possible observations about an object therefore they must select which types of characteristics to focus on. Banning rightly insists that: "Although some archaeologists try to argue that you can collect 'objectively', as in all sciences, we only see something if we are prepared to see it" (Banning, 2000). He then quotes Pirsig:

> According to the doctrine of 'objectivity', . . . We should keep our mind a blank tablet which nature fills for us, and then reason disinterestedly from the facts we observe. But when we stop and think about it disinterestedly,. . . . Where are those facts? What are we going to observe disinterestedly? . . . The right facts, the ones we really need, are not only passive, they are damned elusive, and we are not going to just sit back and "observe" them. We are going to have to be in there looking for them. . . (Pirsig, 1974: 274-75)

Machine learning models and human researchers can approach data analysis from different angles. A researcher starts off by assuming which particular attributes or units of information are most relevant to observe, record, and compare. A researcher is then able to sort examples into meaningful groups based on a pre-defined list of criteria (Banning, 2000). In contrast, a supervised machine learning algorithm first analyses labels by class input examples to empirically identify which attributes are the most relevant for classifying new cases into predefined classes. While even a supervised machine learning algorithm considers only a finite set of features, the difference between a machine learning model and a human expert is that the descriminatory

power of each of the variables considered can be verified objectivelly, enabling selection of the most relevant features that distinguish the training examples from one another.

Applying machine learning to archaeological problems can also widen the degree of complexity in the kinds of questions we can answer about the past. This thesis was initially inspired by Lindahl's paper, which described voids and inclusions in the ceramic paste orienting themselves according to the direction in which clay is pulled during shaping of a vessel.[1](Lindahl & Pikirayi, 2010). The concept of preferred orientation was applied in other studies as well (McGovern, 1989; Berg, 2008; Carmichael, 1986; Blandino, 2003; Rye, 1977; Hill, 1984, 1; Kahl & Ramminger, 2012; Thér, 2015; Ross, Fowler, Shai, Greenfield, & Maeir, 2018). Considering orientation as being simply 'preferred' allowed for making general interpretations regarding shaping techniques.

More complex questions, however, are going to require more precision and rigour in how we collect and analyse data. Given that the orientation of inclusions is a direct outcome of a potter's unique hand movements during shaping, can we identify which pots were shaped by the same person? If handwriting is relatively distinctive to an individual, it is plausible that the same could be said about ceramics made by a particular artisan. In order to explore this further, it was essential to identify a suitable method for extracting, recording, and comparing information related to a preferred orientation of inclusions. All voids and inclusions would need to be identified and a precise relative position of each, along with their angles of orientation,

[1] It has been observed that when a pot is coiled inclusion takes on a horizontal or spiral preferred orientation and when it is shaped using drawing or modelling technique a preferred orientation of inclusions is vertical

would need to be recorded. Deciding how to present and organise such data presented another obstacle. Should it be quantitative in the form of spreadsheets or should it be a graphical representation of an image itself? A new method that could extract and compare relationships between inclusions was required due to the complexity of the information involved. It was in my quest to quantify, record, and analyse this orientation of inclusions that I came across the paper by Shamir discussing pattern recognition techniques for biological image analysis, identifying *wndchrm* as one general purpose machine learning algorithm. (Shamir, Delaney, Orlov, Eckley, & Goldberg, 2010). *Wndchrm* was previously applied to classification problems outside of the biomedical field, as well. For example, it had been successfully used for art style and artist recognition (Shamir & Tarakhovsky, 2012; Shamir, 2012), the study of galaxy images in astronomy (Kuminski, George, Wallin, & Shamir, 2014) and the facial recognition (Shamir, 2011). Given its general purpose algorithmic design, *wndchrm* is likely to be a suitable method for archaeological research as well.

1.2 Academic Contribution

A number of recent studies applied machine learning to classify ceramics by morphological variation. They were meant to explore how machine learning can automate such analysis (Bickler, 2018; Debroutelle, Treuillet, Chetouani, Exbrayat, & Jesset, 2017; Hörr, Brunner, & Brunnett, 2007; Hörr, Lindinger, & Brunnett, 2014, 1; Nguifo, Lagrange, Renaud, & Sallantin, 1997; Makridis & Daras, 2012, 4). A need for increased objectivity, consistency of data analysis, and

speed when dealing with a large volume of ceramic sherds have been identified as key drivers for the push towards adaptation of computerized analysis (Hörr et al., 2014, 1). The proposed method of analysis has not been previously applied to the study of ceramic fabrics and manufacturing techniques. The work presented here was designed to be highly exploratory. The primary goal is to investigate how machine learning can benefit ceramic research in general, identify the challenges of applying *wndchrm*, or a related method of analysis to ceramics, and report on how specific dataset design choices and the training parameters may affect the overall performance of *wndchrm*.

1.3 Thesis Summary

This thesis is composed of four further chapters. In Chapter 2, I describe approaches to archaeological classification and machine learning, while highlighting how *wndchrm* differs from standard systematics used in archaeology. Chapter 3 presents the details of methods used for this project, including the overview of how the *wndchrm* algorithm works, the cultural context of the data, and the methodological considerations related to the sample selection and processing, as well as the dataset design. Chapter 4 presents the results of testing how accurately the *wndchrm* algorithm can distinguish ceramic fabrics, as well as the observations of how different dataset design choices impact accuracy. In the final Chapter 5, I discuss how *wndchrm* can benefit archaeological research and provide practical suggestions for how other researchers in the field can integrate this method into their work.

Chapter 2

Approaches to Archaeological Classification and Machine Learning

2.1 Data Analysis in Archaeology

Archaeological data consists of units of information describing physical materials, or a context in which those materials were created, used, deposited, and recovered (Banning, 2000). The possibilities of the kind of details that can be observed are infinite and researchers have to choose which attributes to focus on. The choice of variables to observe is driven by what is believed to be the most relevant information to the particular questions being explored. In ceramic analysis specifically, three dimensions of artifact variability are typically considered related to the function of objects, their stylistic properties, and their technological characteristics. When looking at functional variability, analyses focus on those attributes that provide the most information about the intended use of a pot (e.g., cooking, storage, display, etc.). In contrast, when studying stylistic differences, consideration will be given to the kinds of attributes that are a matter of cultural or personal preference. And finally, to investigate technological

variability, a researcher would choose to focus on those units of information that have the most potential to tell us about how the ceramics were made. The study presented in this thesis focuses on questions that fall within a domain of technological variability. Heather Lechtman first proposed a concept of technological style noting that a finished artifact is an outcome of all of the decisions made by an artisan during the artifact's manufacturing. (Lechtman, 1977). Some artifact attributes related to ceramic manufacturing processes are noticeable to the naked eye, but others are less obvious as they are embedded in the structure of fired clay, which are often referred to as the ceramic fabric. Ceramic paste and fabric are sometimes used interchangeably. This thesis follows the definition of Rice who makes a distinction between the two. A paste is a processed clay body that is suitable for shaping. A fabric is a term used to describe the paste after it has been fired. (Rice, 1987).

Ceramic fabric can reveal many details about manufacturing processes, such as where the clay was sourced, how it was processed, what kind of temper was added, and how a pot was shaped, smoothed, burnished, dried, and fired. Significant previous work had already been done to identify various relevant attributes of ceramic fabrics to gain insights into the manufacturing processes(Arnold, 1974; Banning, 2000; Berg, 2008; Blandino, 2003; Carmichael, 1986, 1986; Courty, 1994; Hill, 1984, 1; Kahl & Ramminger, 2012; Orton & Hughes, 2013; Rice, 1987; Rye, 1981; Shepard, 1961; Sinopoli, 1991; Ross et al., 2018).

Systematics in Archaeology

Systematics is an area concerned with how to organise, summarise, or group units of information to make comparisons and interpretations. Since the 1950's, a great deal of theoretical discussions have taken place in archaeology regarding how to best arrange and interpret archaeological data. Two major views on this subject have emerged, each gaining its own proponents. The first view was modeled by Albert Spaulding who argued that all attributes found on any given artifact are a direct outcome of intentional choices of the individual(s) who created and used it (Spaulding, 1953, 4). Spaulding's view assumes that there is only a single true way to arrange data and the focus should be on discovering of what the true type is. In contrast, James Ford argued that there are multiple ways in how we can group objects into types, which is driven by the nature of questions being investigated (Ford & Steward, 1954, 1).

The most established approaches to data arrangement typically used in archaeology today are classifications and groupings (Banning, 2000; Dunnell, 1971; Hand, 1997; Adams & Adams, 1991). Furthermore, there are two kinds of classifications (paradigmatic, or taxonomic) (Banning, 2000), and two types of groupings (grounded, or based on central tendency).

In paradigmatic classifications, categories are non-hierarchical and non-weighted. In contrast, taxonomic classifications employ weighted and hierarchical categories featuring main categories, sub-categories, and sub-subcategories. The standard classification starts with a formulation of conditions that will need to be met to qualify for class membership. All rules defined for class membership are strict, meaning they must be fully satisfied for a particular classifica-

tion decision to be valid. These class conditions are also considered sufficient, meaning if a list of specified requirements is fully met, there is no ambiguity about class membership. The rules of classification are independent of the actual collection of material and could potentially result in some classes being 'empty', with no real examples (e.g., there are red-slipped pots with shell temper, red-slipped pots without shell temper, plain pots with shell temper, but no plain pots without shell temper).

In contrast to classifications, a grouping arrangement is driven by a collection of real objects. Observed variabilities between actual objects are considered for establishing conditions of how to sort and group data. Grouping rules can evolve when new examples are added or removed from a collection. Conditions for grouping objects are ambiguous because they are neither strict or sufficient, leaving room for flexibility to manipulate rules of arrangement as needed to fit a particular collection. While groupings can be distinctive, they can also be less defined simply clustering similar objects together. A bounded grouping is a contextual kind of data arrangement based on boundaries that are extrinsic to the object itself. Bounded grouping is premised on the idea that objects are chosen to be compared as part of the same collection due to their shared context(Banning, 2000). For example, artifacts that come from the same location, or represent the same occupational period would be considered together. The central tendency method [1] places more focus on intrinsic attributes of objects, while the contextual information is given less attention. In this case, artifacts are compared based on the central tendency of each group

[1] Central tendency can be easily measured by calculating the mean value when working with data derived from the interval or ratio scales of measurements (Banning, 2000)

and the distance measurements. The distance measure can be defined as dissimilarities between items in a multi-dimensional space (Banning, 2000).

2.2 The Field of Machine Learning

Machine learning is an interdisciplinary area of computer science and applied statistics, overlapping with related disciplines including mathematics, physics, engineering, computer science, cognitive science, computational neuroscience, and economics. Arthur Samuel is widely accepted as the first to establish the term machine learning in 1959, when he attempted to program a computer to play checkers (Samuel, 1959). Samuel defined machine learning as giving computers the ability to learn without being explicitly programmed. Machine learning algorithms learn to complete tasks or make predictions when provided with some input information or experiences that can be used to inform future interpretations when the model is presented with new information. (Kohavi & Provost, 1998). There are three main approaches to machine learning supervised learning, unsupervised learning, and reinforcement learning (Mackay, 2003; Bishop, 2006; Hastie, Tibshirani, & Friedman, 2009; Shai & Shai, 2014).

Supervised Learning

Supervised learning is commonly used for classification and regression problems. Training is an essential part of the supervised method. In the case of *wndchrm* specifically, the algorithm searches for the most relevant attributes by taking into account how training examples

are assigned into their respective classes. Input data in this particular case consists of training examples and details about their assigned class membership. Output information is in the form of a prediction, assigning new examples into one of the pre-defined classes.

An example of a problem well suited for supervised machine learning approach is to classify images of cats, dogs and mice. First, labelled examples of each kind of animal will be provided to train a model. To test how well the trained model can classify new pictures, a researcher will present unlabelled images to the trained classifier to be categorised into one of the three classes. In this particular example, the predicted output is qualitative. Another variation of a prediction is ordering data categorically. For example, samples can be ordered into categories such as Small, Medium, and Large, or as First, Second, or Last. Regression is an example of yet another problem suitable for a supervised approach, but with a quantitative prediction output. An example of regression would be to predict the exact weather for tomorrow by looking at the historical weather data.

Unsupervised Learning

In contrast to the supervised approach, when unsupervised learning is applied the input data is uncategorised. In unsupervised machine learning, interpretations and predictions are based on exploring patterns in unlabelled data. Clustering is an example of the unsupervised approach, in which input examples are grouped into clusters rather than distinct labelled classes. Mapping of input training examples based on their distances from one another results in clustering of the most similar images closest together. A common example of this method is k-means clustering

algorithm (Theobald, 2017).

Reinforcement Learning

Unlike supervised and unsupervised models that reach their full capacity once trained, the reinforcement learning model continues learning and improving indefinitely, as it becomes exposed to additional information and experiences. Reinforcement learning is the least likely to be used with archaeological data, however, as one of the three major approaches to machine learning, it deserves a mention.

In the case of reinforcement learning, a model continues to learn and improve as it is exposed to new experiences. For example, reinforcement learning can be used to train an algorithm to play chess. Training occurs by signalling failures and successes, often through a reward. In this case, one could 'reward' successes, such as taking opponents piece, or not losing its own piece as +1, and losing a piece as -1. The more the algorithm 'practices' playing the game, the more it learns how to make better moves in the game. Similarly, reinforcement learning can be applied in robotics, where a robot must learn to move around a landscape effectively and avoid obstacles. Reinforcement learning is ongoing, meaning an algorithm could theoretically continue to become 'smarter' indefinitely as it is exposed to new data and experiences.

2.3 Aligning Machine Learning and Archaeological Terminologies

The terms class and classification are not defined in the same ways in supervised machine learning as they are in archaeology. Machine learning fits neither the definition used in archaeology to describe classification nor the grouping methods for data compilation. While it may resemble some elements of the classification approach and some of the grouping, machine learning approaches are distinct from either.

	Classification	Grouping	Wndchrm
Conditions:	- necessary and sufficient - independent of the data - permanent	- central tendencies - driven by the uncategorised data - can change	- weighted tendencies - driven by the categorised data - permanent
Membership:	- definitive - empty classes are possible - mutually exclusive	- ambiguous - all groups have members	- probabilistic - mutually exclusive - all groups have members from training examples, but could not have any in from the new uncategorised examples.
Variation:	- paradigmatic classification - taxonomy	- bounded - attribute association	- nearest neighbour - nearest distance to a class mean (WND5)

Figure 2.1: Conventional approaches of data arrangements used in archaeological compared to *wndchrm's* approach.

Returning to Samule's definition of machine learning discussed earlier, a computer is not explicitly programmed to perform a function, or in the case of supervised learning, to merely sort (or classify) data using some pre-defined rules. After learning and analysing training examples,

an algorithm is programmed to establish rules for classification. Once the rules are established, a model is then ready to use those rules to classify new uncategorized examples into one of the defined classes. In a standard classification arrangement of data found in archaeology, the conditions of class membership are strictly defined first, independent of actual data being classified. Membership in a typical class as described in archaeology is definitive, meaning there is no ambiguity and any given object can only be classified in one of the specified classes.

In contrast, grouping is descriptive and based on the actual data being interpreted. Grouping arrangements don't have formal, strict, and permanent rules, therefore an object can be assigned into different groups, or group membership can change, especially as examples are added or removed from the mix. In contrast, *wndchrm*, which is an example of supervised machine learning, uses a probabilistic approach. While there can only be a single choice when assigning a new example to a class, it is not definitive in the same way as the standard classification discussed, or as flexible as the grouping method. Instead, it is probabilistic and a decision is made based on quantifying the most probable class membership, by identifying a class with shortest distance to an unlabelled example. [2] Once an algorithm is trained, the rules are formalised and are no longer altered when supervised learning is used. A researcher can choose to train the original algorithm again using a different set of training examples. Figure 2.1 compared the classification approaches, the grouping approaches and the *wndchrm* method discussed, by outlining some of the key characteristics of each.

[2] Quantifying the probability of an object belonging to a particular class is one of the strengths of using machine learning over a more conventional classification.

2.4 Practical Challenges of Application

A good algorithmic design is crucial for developing an effective machine learning model. However, an algorithm is unlikely to reach its full potential if provided with a dataset of poor quality. Optimal dataset design requires some thought and experimentation. Data can come in many forms. It can be qualitative, quantitative, or graphical. It may be partial, or incomplete. The original population size represented by a given sample may be unknown. In some cases, details about training examples themselves may be hypothetical rather than confirmed.

In the case of archaeology especially, the data can be very complex and all of the above challenges are common. A significant experimental and theoretical work, therefore, is required to establish best practices for collecting, documenting, and selecting archaeological data for machine learning analysis. Standard practices will ensure comparability of datasets and enable comparison of data derived from different projects, archaeological sites, and contexts.

Chapter 3

Methods

3.1 Analytical Objectives

At present there are no established practices or procedural recommendations in place for using supervised machine learning to classify pottery. This study evaluates the ability of the *wndchrm* algorithm to accurately identify ceramic fabrics by shaping technique and paste recipes, with an objective to provide feedback for future work in this area, especially as it relates to:

a. Sample selection.

b. Physical samples processing.

c. Digital data collection methods.

d. Dataset design.

e. Sample size and accuracy considerations.

f. Parameters setting for training and testing.

g. Strength and weaknesses of *wndchrm* for ceramic studies.

3.2 Wndchrm Model

Method Summary

Wndchrm is an example of supervised machine learning approach. It was developed in the United States by the Goldberg Group at the National Institute on Aging in Baltimore for classification of biomedical images (Shamir et al., 2008; Shamir et al., 2010; Orlov et al., 2008). *Wndchrm* is an open source utility available to be freely used by anyone. It consists of a collection of algorithms designed to extract information from training examples, identify and prioritise attributes that are most discriminative for assigning training images into classes, and classify new data. The entire process employed by *wndchrm* consists of three key steps. First, the feature extraction algorithms are used to transform graphical contents of pictures into numerical values. The extracted features are then given weights and ranked by how meaningful each feature is to an assignment of training examples into classes. To classify new images the *Weighted Nearest Distance* approach is used by the trained *wndchrm* classifier. (Orlov et al., 2008; Shamir et al., 2008).

3.2.1 Implementation

Wndchrm is written in C++ and uses *libimfit* libraries to compile all algorithms making it easy to integrate with other software tools and packages. *Wndchrm* does not have a graphical user interface instead it has the basic ability to execute commands and programs in a command line utility

(Shamir et al., 2008). The list of commands specific to using *wndchrm* is well documented and Shamir et al. (2008) provides details and instructions about how to use *wndchrm*.

The required skills for using *wndchrm* can be learned fairly quickly, and programming knowledge is not necessary. The *Wndchrm* commands consist of two categories of actions. The *wndchrm train* command is used to extract image features and document graphical information numerically and the *wndchrm test*, or the *wndchrm classify* commands are used to assign weights to features based on their ability to discriminate between classes. Because *wndchrm's* train command's primary function is to extract features, the same script is executed to compute features from both training images, as well as unlabelled new data. *Wndchrm's* implementation was designed to enable researchers to experiment with different testing parameters to obtain best results without having to extract features every time, as it would be very time-consuming. Both feature selection and class prediction take place when the *wndchrm test*, or *wndchrm classify* command is executed.

In the example shown in Table 3.1, *wndchrm* is instructed to extract a large set of features (-l) as well as a colour feature (-c), followed by a link to a directory where images are stored. A second path is provided indicating the destination where an output file with extracted information could be created.

In the example shown in Table 3.2, *wndchrm* is instructed to select the most relevant features based on how the training examples were assigned into classes. An address to a feature file (.fit) containing information about the extracted features for all images and classes is given. *Wndchrm* is being asked to use only 20% of highest ranked features out of all extracted ones

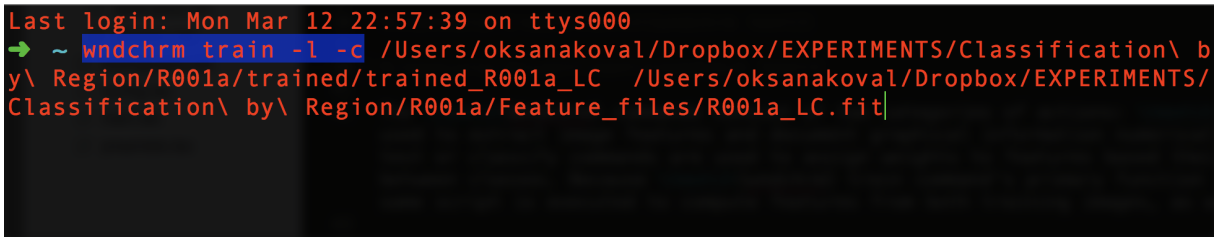


Figure 3.1: An example of a command to extract features.

(-f.20) to classify test examples. A path to a directory is also provided to create a (.html) file to document results of an experiment.

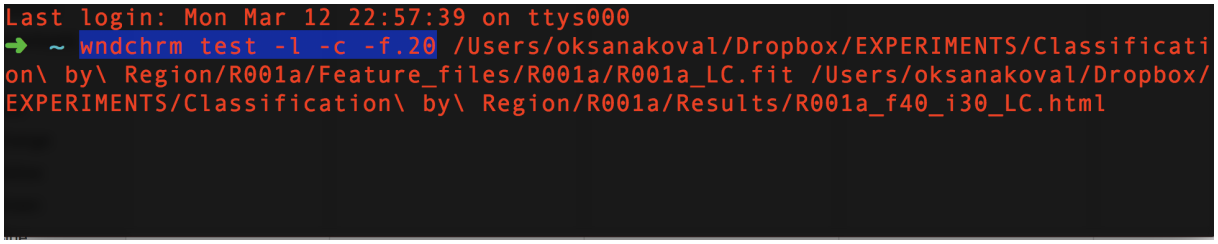


Figure 3.2: An example of a command to assign weights to features and cross validate classifier.

Input Data

Input data consists of image examples in .tiff file format. Training examples are 'labelled' by using directories. Images from the same class are stored together in a separate subdirectory. Splitting training images into subdirectories is a way in which *wndchrm* algorithm recognises a class of images.

Feature Extraction

When images are processed, a separate .sig file is created for each image, as shown in Figure 3.4. After all of the image files in a given directory are processed, a .fit summary file is also created to summarise information about the extracted feature values for all image files in the dataset.

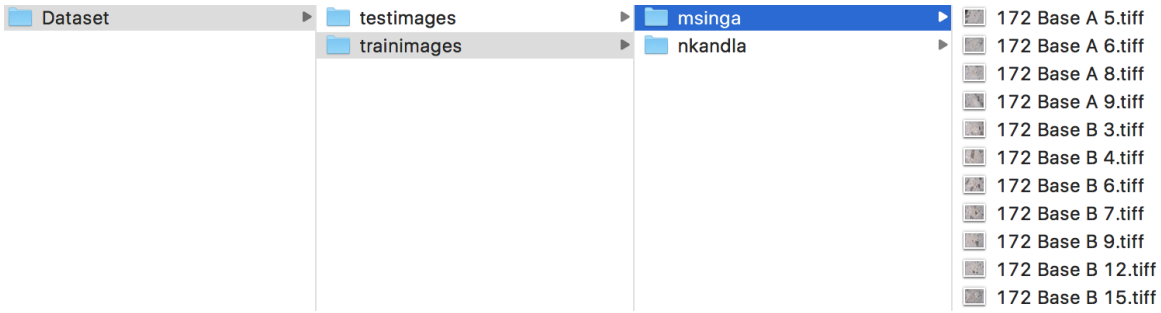


Figure 3.3: An example of a training dataset consisting of two classes.

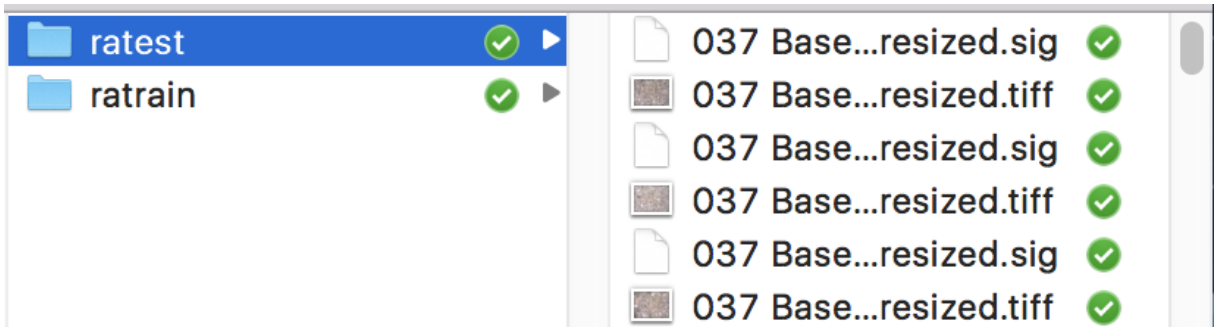


Figure 3.4: An example of a directory with .sig files.

Cross-Validation

The *test* command is used to cross-validate accuracy of a classifier. One has a choice to either identify which specific examples should be used for training or to have the training and testing examples be chosen randomly. When the *wndchrm test* command is executed, test images are randomly pulled from a class-specific directory to be used as test examples. *Wndchrm* is instructed to disregard information about a class membership of these images. This option allows a researcher to specify a proportion of images to be used for training and testing to cross-validate accuracy. The *test* option can be especially advantageous for small datasets, enabling a higher number of examples to be used for training and as little as one image can be used for testing (Shamir et al., 2010). This can be accomplished by running a script with the (-n) parameter specified, requesting multiple random splits, and alternating different images used for training and testing. When multiple splits are used as part of a single experiment, the results of all splits are taken into account to determine the accuracy of the classification. Accuracy is calcualted as a percentage of all images classified correctly out of total number of images tested.

An alternative option for cross-validation is to specify which particular images are to be used for training and testing. To accomplish this, training and testing images would need to be placed in separate directories and the *wndchrm classify* command should be used instead of the *wndchrm test* command. The only difference between these two commands is that *wndchrm classify* allows for selecting which specific files should be used as test examples.

Classifying New Data

In order to use *wndchrm* to classify new samples of an unknown class against a training dataset, a *wndchrm train* command is used first to extract image features from unclassified images. The *wndchrm classify* command would then be used to classify new data.

3.2.2 Image Content Descriptors

Pattern recognition classifiers can be designed to take into account a wide range of image features. Some classifiers are tailored for a very particular classification problem, while others can tackle a wider range of images and datasets. The task-specific classifiers typically use a smaller number of features, heavily informed by previous experiments, as well as, domain knowledge of a human expert (Orlov, Johnston, Macura, Shamir, & Goldberg, 2007). In contrast, multi-purpose classifiers are typically programmed to extract a greater variety and number of image features. Taking into account a greater variety of features makes such algorithms more suitable for classifying a broad variety of images (Orlov et al., 2008).

Wndchrm is a multi-purpose classifier that extracts up to 4,059 image features. The collection of algorithms implemented in *wndchrm* computes the long list of image features by transforming graphical information into numerical values.

Wndchrm takes into account four categories of features polynomial decompositions, high contrast features, pixel statistics, and textures (Orlov et al., 2008). Polynomials are the closest to approximating contents of an actual image. Contrast features, such as edges and objects, include

statistics about the numbers of objects, their spatial distribution, size, and shape. Pixel statistics are based on a distribution of pixel intensities, in a form of histograms and moments for an original image as well for a transformed image. Fourier, Wavelet, and Chebyshev transforms are used for this [1]. The image content description algorithms used by *wndchrm* are well documented in the literature and can be easily referenced for additional information. (Shamir et al., 2008). They include:

- Radon Transform Features (Lim, 1990)
- Chebyshev Statistics (Gradshtein & Ryzhik, 1994)
- Gabor Filters(Gabor, 1994; Gregorescu, Petkov, & Kruizinga, 2002)
- Multi-Scale Histograms (Hadjidementriou, Grossberg, & Nayar, 2001)
- Fist 4 Moments
- Tamura texture Features (Tamura, Mori, & Yamavaki, 1978)
- Edge Statistics Features (Prewitt, 1970; Murphy, Velliste, Yao, & Porreca, 2001)
- Object Statistics (Otsu, 1979; Gray, 1971)
- Zernike Features (Teague, 1980; Murphy et al., 2001)
- Haralick Features (Haralick, Shanmugam, & Dinstein, 1973; Murphy et al., 2001)

[1] Transform of an image is a variation of an original image, distorted or significantly changed, with the same pixels being represented differently, in different patterns following different rules.

- Chebyshev-Fourier Features (Orlov et al., 2007)

3.2.3 Feature Selection

Particular types of image content descriptors that would be useful are highly variable by context. When multi-purpose classifiers extract large sets of features, only a small subset of them will be relevant to any given classification problem.

Taking into account an entire large set of features of various relevance can constitute a problem, often referred to in machine learning as *noise* (Orlov et al., 2008; Shamir et al., 2010; Shamir et al., 2008). Noise describes a situation when dubious features are considered alongside those with more discriminative capacity, reducing the overall weight given more relevant features. For this reason, taking into account all features is likely to have a negative impact on classification accuracy. Feature selection process addresses this issue, because it identifies a smaller subset of extracted features with most discriminative properties and disregards all others from being considered by a classifier.

Pattern recognition algorithms typically use either a wrapping or a filtering approach to select features. (Shamir et al., 2010). Wrapping method selects a set of the most relevant features by trying out different subsets of features in a classifier to test which set of features performs best when used to classify test images. *Wndchrm* does not test features in a classifier, instead it employs an approach known as filtering. *Wndchrm* computes fisher scores for all extracted features to identify those features found in training examples that are most informative to how

they are assigned to classes. Every approach has its downsides and filtering has an increased risk of selecting multiple related features [2], which may become over-represented resulting in noise rather than improving the effectiveness of a classifier (Shamir 2010). Paying attention to what features are being selected and their relative fisher scores is a critical step for monitoring this concern.

Fisher Score

Wndchrm uses a statistical method known as filtering to select features based on their ability to discriminate classes (Shamir et al., 2008; Orlov et al., 2008). A fisher score is calculated for each feature. A fisher score can be defined as a ratio of the variance in the feature value between classes to its variance within classes. Those features with the lowest fisher scores are discarded. This stage of analysis is known as the feature selection because a list of features is greatly reduced as less relevant features are discarded.

Fisher scores act as weights for ordering features by their discriminative significance. Those features with the lowest within class variance and the highest variance between classes are most discriminative and rated highest.

In probability theory, variance measures how far a set of values are spread out from their average value. To calculate the variance for a population:

1. Calculate the mean of the population. [3]

[2] A group of features may be similar within the same general category of features.

[3] The formula to find a population mean is: $\mu = (\sum * X)/N$, where: $\sum$ means "the sum of", X represents all the individual items in the group and N is the total number of items in the group.

2. Subtract the mean from each number and square the result. [4]

3. Calculate average of the squared differences.

During the training stage, image features were extracted from each image with a separate numerical value computed to describe different image content. As described below, *Within Class Mean* is calculated for each class based on feature values extracted from all training images in a class. *Mean Class Values* are then used to calculate the *Pooled Class Mean*, that is subsequently used to calculate the *Variance of Class Means*.

The *fisher score* is a ratio of the variance in the feature value between classes (*Variance of the Class Means*) to the variance within classes (*Mean of Within Class Variances*). The fisher score is calculated with weights being assigned to each feature based on their discriminative power assessed through consideration of how training examples are assigned into classes.

W_f, the weight assigned to the feature $_f$, is calculated as follows:

$$W_f = \frac{\sum_{c=1}^{N}(\overline{T_f} - \overline{T_f, c})}{\sum_{c=1}^{N} \sigma_{f,c}^2}, \tag{3.1}$$

where $\overline{T_f}$ is the mean of the feature f values of all image examples in a dataset, $T_{f,c}$ is the mean of the values of feature f in certain class c, and $\sigma_{f,c}^2$ represents the variance of feature f in the image examples of class c.

[4] The result is squared to make negative numbers postive, as it is the distance from the mean which is important rather then negative or positive nature of value

Number of Features

Wndchrm allows for flexibility to experiment with using different numbers of features. When executing *wndchrm train* or *wndchrm classify* commands, an additional parameter *f* can be added to specify what percentage of all extracted features should be kept and what percentage should be discarded. If *f* parameter is left unspecified, *wndchrm* will by default use top 15 percent of most discriminative features.

Reducing or increasing a ratio of total features to be used can impact the accuracy of classification. Therefore, running multiple experiments while varying the *f* parameters enables a researcher to identify a preferred number of selected features that should be used for a particular classification problem.

One of the limitations of *wndchrm* implementation is that it does not allow for the selective elimination of specific features. Performance of a trained classifier could be negatively impacted if multiple selected features represent the same general group of features and are very similar. This can create noise and prevent other more diverse features with slightly lower fisher scores from being considered. The values of weights assigned to selected features used in classification can be found in the *wndchrm's* report file. This file should be carefully reviewed to note what features are represented in classification.

3.2.4 Cross-Validation and Classification

After determining which features are most relevant, the *wndchrm's* trained model is then tested by being asked to classify new unlabeled data. This stage is known as cross-validation. The images used for cross-validation are commonly referred to as control or test examples. Class assignment of control images is known to the researcher but is not known to the trained classifier being tested. Classification accuracy is measured by dividing the number of test images that were classified correctly by the total number of images that the algorithm attempted to classify (Shamir et al., 2010).

K-Nearest Neighbour

Nearest neighbour is a common approach used by pattern recognition classifiers. The central principle of the nearest neighbour approach is that an algorithm looks for a training example(s) that has the most similar feature values to that of a test example. Consider a problem where images are classified as either a strawberry or a banana. For this demonstration, two types of features are considered weight (x) and colour (y). Eight training examples are used per class and each example is plotted as a point in Euclidean space after taking into account weight and colour values. Training samples for the banana class are shown in yellow and those for the strawberry class are shown in red 3.5.

First, the Euclidean distance D between examples is determined, which can be defined as a straight line between two points in the Euclidean space and is calculated as follows:

Figure 3.5: Training examples plotted as points in Euclidean space.

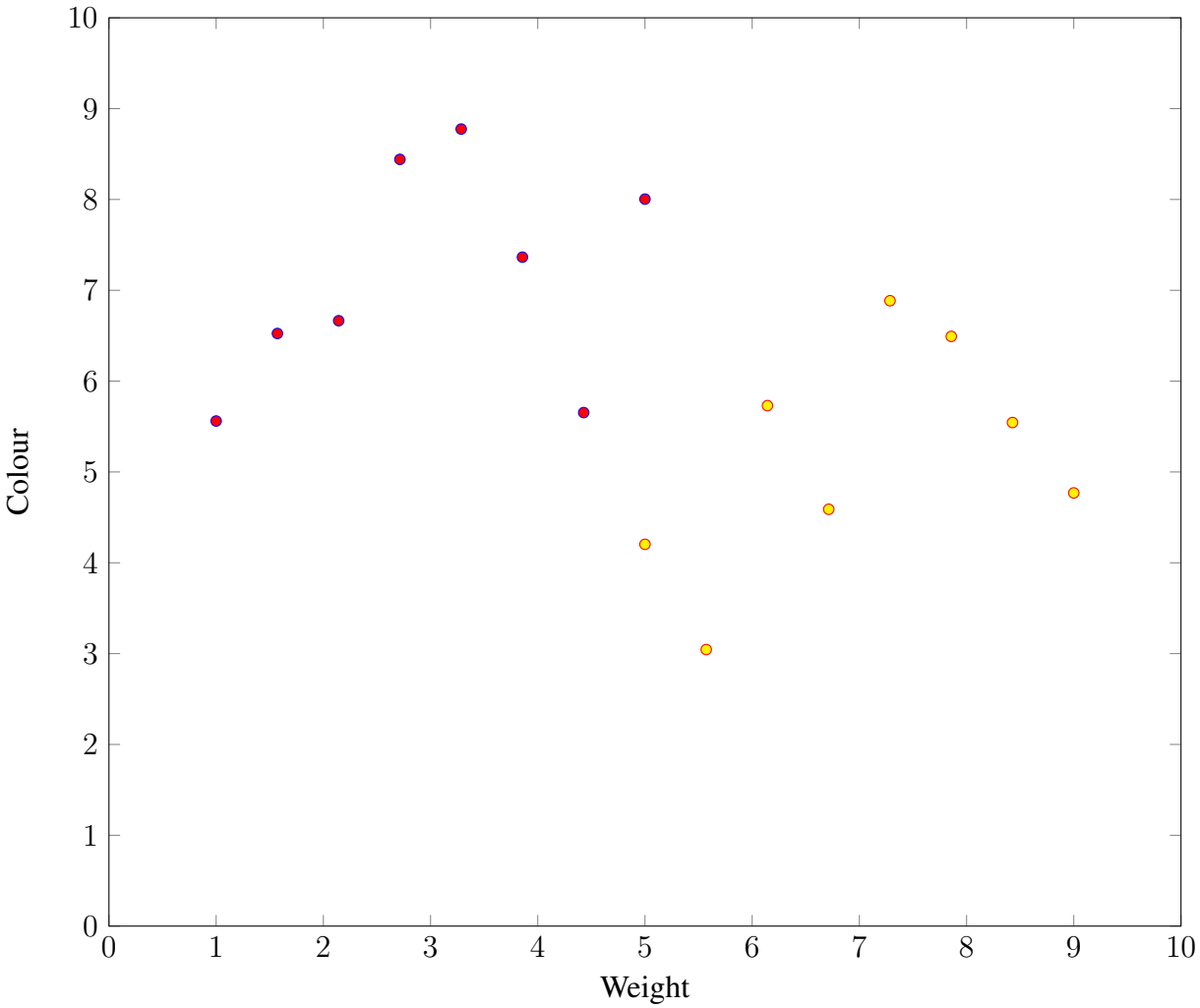
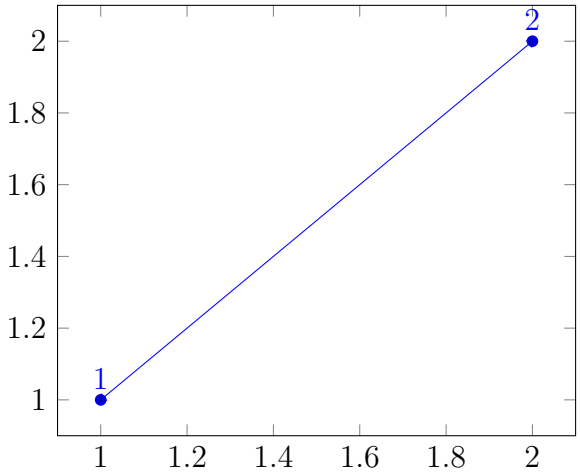


Figure 3.6: Graphical representation of distance between two examples.



$$D = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}, \tag{3.2}$$

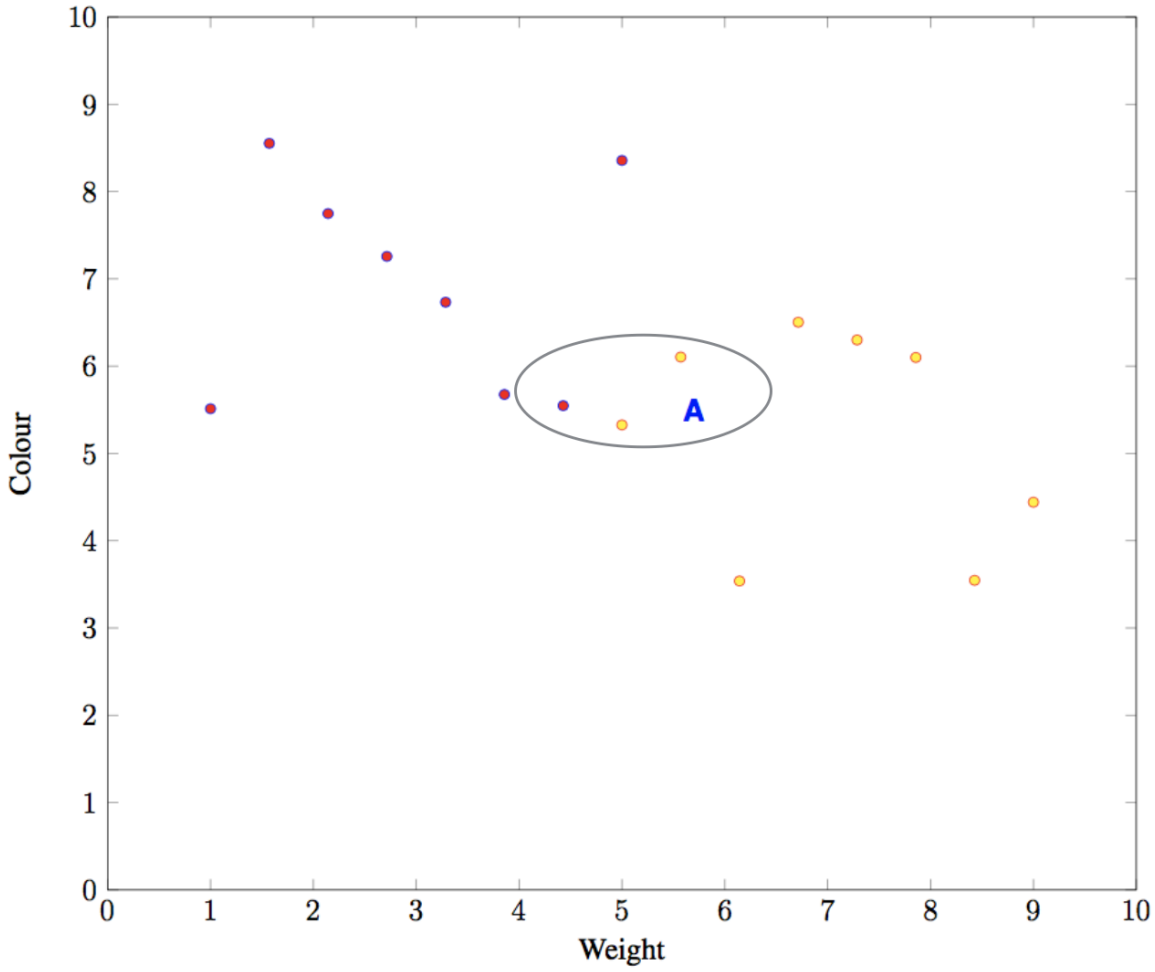
The above calculation assumes two feature values are being taken into account (x, y). The same approach works with more than two features in multi-dimensional space. For example, one could consider four types of features to classify images as either bananas or strawberries, such as weight (x), colour (y), texture (b), and length (b). In this case, the Euclidean distance D would be calculated as follows:

$$D = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (a_1 - a_2)^2 + ((b_1 - b_2)^2}, \tag{3.3}$$

In K-nearest neighbour, K value determines how many closest training examples are taken into account. For example, if $K = 1$, than a unknown image will be assigned to a class of a single closest in distance training example. If K is greater than one, then a new example will

be assigned using the majority vote of its closest neighbours. Odd K values work best, because if K is an even number, and there is a tie, than deciding between classes that are tied would be random in nature.

Figure 3.7: An example of K-nearest neighbour classification, where $k = 3$.



In the example shown in Figure 3.7, a new image A, is to be classified by considering three

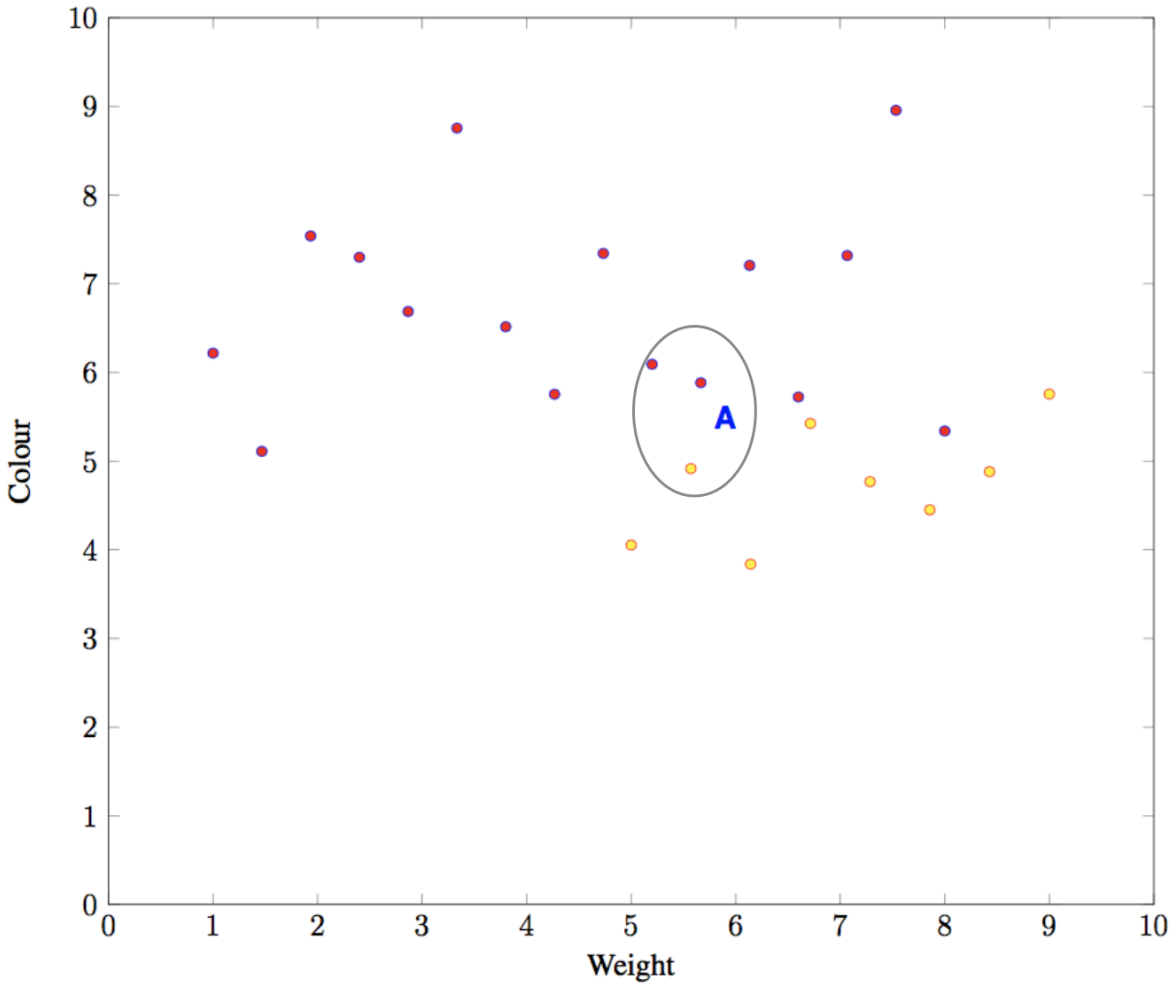
training images ($K = 3$) with the shortest Euclidean distance to the unlabeled image A. The diagram shows two of the closest training examples are classified as a banana in yellow and one as a strawberry illustrated in red. By the rule of the majority vote of K-nearest neighbour, the decision, in this case, was to classify the new image as a banana.

A class represented with more training examples is likely to dominate. Given the majority vote approach, using the same number of training examples for each class is advisable to avoid unbalanced training. Figure 3.8 below illustrates the same classification problem as 3.7, but this time demonstrating a negative impact of the unbalanced training. In this case, 16 training examples were used for the strawberry class and 8 training examples were used for the banana class. Similarly to the previous example, the K value of three is used ($k = 3$), however in this case, the classification result is incorrect caused by the strawberry examples being overrepresented.

It should have now become apparent, that how images are plotted in Euclidean space and values of their feature vectors are directly affected by the kinds of features being used to plot each example. If features that are not discriminative (very similar for all example belonging to different classes) were to be used, the accuracy would be negatively impacted. For this reason, the feature selection step is the most critical part of these analyses designed to eliminate features that are either too redundant and overrepresented as a group or are irrelevant and non-discriminative.

Wndchrm uses a variation of the standard K-nearest neighbour approach described known as the probabilistic nearest neighbour. Basic K-nearest neighbour gives equal weights to all selected features when calculating the distance. In contrast, the probabilistic nearest neigh-

Figure 3.8: An example of unbalanced training.



bour takes into account weighted values assigned to each feature to plot data and calculate distance. *Wndchrm* offers a choice of using either the Weighted Nearest Neighbour *(WNN)* or the Weighted Nearest Distance *(WND)*.

Weighted Nearest Neighbour

(WNN) is a variation of the nearest neighbour that accounts for weights assigned to each selected feature when plotting data and calculating distance to nearest examples. In order to use this approach with *wndchrm*, a parameter (-w) should be added when executing *wndchrm test* or *wndchrm classify* commands. *Wndchrm* classifier will then search for the nearest image example with the shortest distance to a new image being classified. A class to which the nearest training example belongs will be a classification result of a new image.

In Weighted Nearest Neighbour approach, the distance of a feature vector x from a certain class c is the shortest Euclidean distance between the new example and any training example of that class are calculated using the following Equation:

$$d_{x,c} = \min_{t \in T_c} \sum_{f=1}^{|x|} W_f (x_f - t_f)^2, \tag{3.4}$$

where T_c is the training set of class c, t is a feature vector from $T_c|x|$ is the length of the feature vector x_i, x_f is the value of image feature f vector x_i, and W_f is the Fisher score of feature f.

Weighted Nearest Distance *(WND5)*

If the (-w) parameter is not specified, *wndchrm* employs the Weighted Neighbour Distance *(WND)*, which evaluates the probability distribution of new images belonging to each of all possible classes. Unlike the *WNN* approach that assigns a new image to the same class as the nearest training image, *WND* considers the distance to each class as a whole, which is based on the mean features values of all the training images. The nearest distance between the new example and a class is the distance from a new image to all the training images in each class is considered.

The distance between one image and a certain class c is calculated as follows:

$$d(x,c) = \frac{\sum_{t \in T_c} [\sum_{f=1}^{|x|} W_f^2 (x_f - t_f)^2]^p}{|T_c|}, \tag{3.5}$$

where T is the training set, T_c is the training set of class c, t is a feature vector from T_c, $|x|$ is the length of the feature vector x, x_f is the value of image feature f, W_f is the Fisher score of feature f, $|T_c|$ is the number of training examples of class c, $d(x,c)$ is the computed distance from a given example x to class c, and p is the exponent, which is set to -5.

3.3 Experimental Data and Cultural Context

The experimental data used in this case study consists of the modern day ceramic vessels collected by Kent Fowler. Fowler has visited several Zulu communities in South Africa and docu-

mented details related to the local ceramic manufacturing practices observing the full cycle of the production process starting from clay acquisition to final product. The videos, field notes, and publications from his fieldwork provide the necessary insights into the ceramic manufacturing practices, the different types of vessels functions, as well as more general information related to the cultural and social context. (K.D. Fowler, 2006, 2008, 2011)

Most of the Zulu ceramic data used in the experiments discussed in the next chapter, came from two regions Nkandla and Msinga. Nkandla pottery came from the community known as Magwaza. Msinga collection represents two communities Mabaso and Mchuno.

3.3.1 Morphological and Functional Classification

Zulu ceramics can be broken down into several distinct groups according to their morphology and intended functions. General function categories include food preparation, serving and drinking, storage and transport, and medicinal and ritual uses. Figure 3.10 illustrates the common morphological types and their functions.

The ceramic data used for this study include three Umancishana (13), six Izinkamba[5](12), and one Ingcazi (15).[6] The izinkamba vessels are primarily used for beer serving and drinking. The umancishana are the smallest serving vessels roughly 15 cm in height. The umanchisana are least frequently used for drinking, but are used ceremonially for serving beer to the ancestors. This is likely due to its smaller size compared to the other drinking vessels. Izingcazi[7] can

[5] Ukhamba is a singular form of Izinkamba.
[6] Several other vessels were also processed and are listed later in this chapter, however, these samples were only used in a limited number of experiments.
[7] Izingcazi is a plural form of Ingcazi.

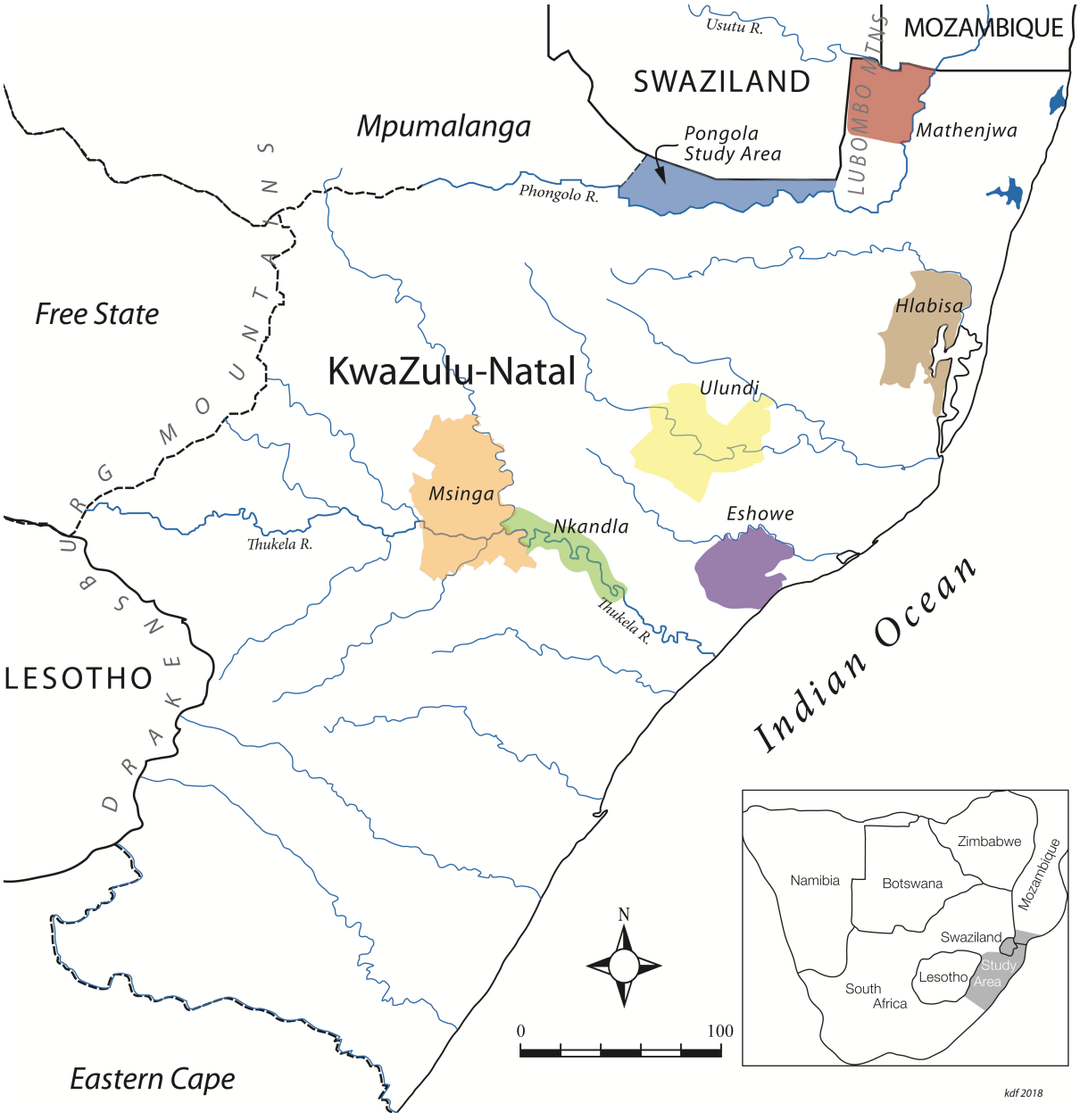


Figure 3.9: Regional map of Zulu communities in South Africa. Map provided courtesy of Kent D. Fowler.

vary in size anywhere between 14 to 43 cm and are used to transport water or beer. While the body part of the incgazi vessel is similar to ukhamba and umanchisana, it includes a long everted neck. The morphological variation of having a neck, makes incgazi more suitable for its function of transporting liquids.(K.D. Fowler, 2006)

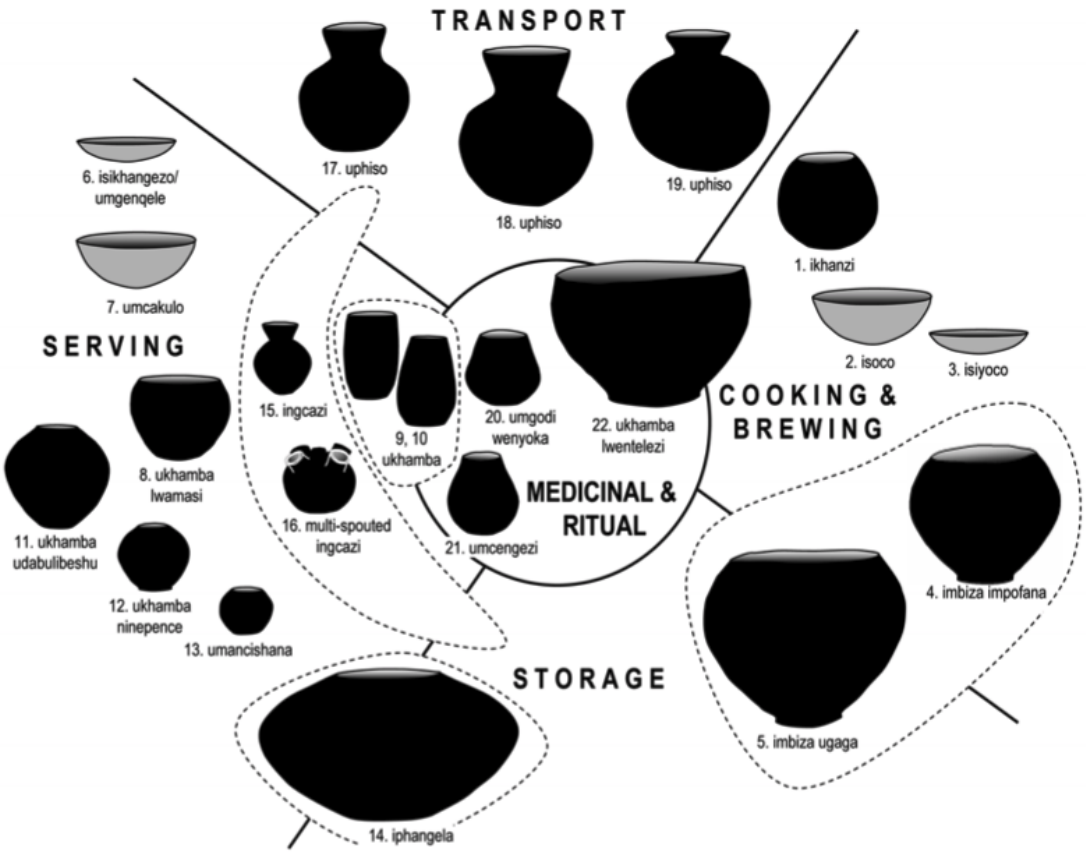


Figure 3.10: Classification of Zulu vessels by form and function. (K.D. Fowler, 2006)

3.3.2 Clay Processing and Paste Recipes Represented

Local Clays

The potters from Nkandla source two types of clays, which can be easily distinguished as red and black variety. In the Msinga region, the clay to make pots is sourced from dried out streams. The Msinga clay is the reddish-brown calcareous clay, that is fairly fine in texture. The clays in the Msinga region are high in organic content and contain rock inclusions such as quartz, feldspars, metamorphic limestones and granites, and calcite granules. These clays primarily consist of illite and kaolinite. In Mabaso, the potters source two types of clays to make ceramics. One is the coarser clay and the second is the finer sandy clay. The clays used by the Mchuno potters are very similar in mineralogical composition to those used by the Mabaso potters. The clays used by the Mchuno community derive from four different locations. Two of the Mchuno sourced clays are finer grained variety and the other two are coarser grained. Fowler provides details of how clays were sourced, processed, and tempered to prepare paste, indicating similarities and some differences between regions and functional types of vessels (K. Fowler, Middleton, & Fayek, 2017). Below is a brief overview of how Zulu potters prepare thier paste.

Clay Processing

Once clay is acquired, the potters dry it for several days. Processing steps involves turning clay into a fine powder. Large inclusions are first removed by hand. Some regional variation was observed in how clay is processed to achieve a finer texture. The Magwaza potters from

Nnkandla and the Mchuno potters from Msinga, use grindstones in a rolling motion to grind clay into a fine powder. The Mabaso potters from Msinga pound their clays with a stick instead. The grinding technique specifically was observed to produce a slightly finer clay.

Tempering

Depending on the intended function of a vessel, different temper material is used to mix the paste. To manufacture a drinking vessel, the potters from Nkandla mix equal proportions of the processed black and red clays to achieve the desired plasticity. The potters from Msinga also used secondary clays as tempering material when making drinking vessels. After removing coarse inclusions and processing the primary clay into a fine powder, they used a secondary clay as tempering material to make a paste. The proportion of added secondary clay and resulting coarseness of temper varies depending on the type of vessel being made. To achieve the finer paste required for making drinking vessels, a lesser proportion of secondary clay is added in comparison to the other types of vessels.

3.3.3 Shaping Techniques Represented

Fowler (K.D. Fowler, 2006, 2008, 2011) reports that all the potters were highly consistent in how they shaped vessels. The bottom of a pot was made first out of a lump of clay using what is known as the slab shaping technique, where a piece of prepared clay is shaped into a slab disk. The outer edge of this disk is drawn up minimally to make a lip. The slab is about 1 cm thick and is the thickest part of the vessel. The rest of the pot is built in sections using the coiling

technique, with a first coil being attached to an interior side of a lip followed by additional coils placed over each other to build a vessel wall. Coils are made by rolling clay vertically with two hands. In smaller vessels, coils are roughly 10-12 cm long and about 1.5 cm thick.

The Magwaza potters from Nkandla and the Mabaso potters from Msinga add coils and smooth vessels gradually, building it in three sections. The Mchuno potters, on the other hand, shape the entire form first and then use the maize cob to smooth a pot. To build an upper section of a pot, coils are gradually reduced in size. After a pot is smoothed and coils are well joined, the very top of the vessel is cut with a knife into a circular opening. Once dried, a vessel is decorated, using a combination of incision, excision, impression appliqué, and burnishing. After decorating, a pot is dried further (for up to a week) and then fired. This shaping method applies to all Zulu ceramics. Depending on the size of the vessel, the angle of the lower portion of a wall will vary. To shape the incgazi vessel with an elongated neck, two or three additional coils are added in the opposite direction to widen the wall at the end.

3.4 Methodology

3.4.1 Sample Selection

The key intention when selecting vessels for this study was to choose a combination of pots that are comparable in shape and size to control for data consistency. The best way to optimise for multiple classification problems was also taken into account. Therefore, the pots for the exper-

iments came from several different communities to enable classification experiments based on regional differences in raw materials and paste preparation process. Additionally, both bases and walls were included in the analysis to explore whether the two can be distinguished by the differences in how they were shaped. And lastly, several potters were also chosen to attempt classifying ceramics made by different potters.

3.4.2 Pottery Processing

The ceramic material had to be processed physically in a lab. Most of the vessels used were cut in half vertically, using a wet saw. The photographs of the ceramic samples in the collection that were processed for this study are shown in Figure 3.11.

All the samples were lightly polished. A quick 10 minute polishing of each thick section made a notable difference. Figure 3.12 shows the difference in image quality between polished and unpolished scanned sections.

3.4.3 Image Acquisition

Two methods for collecting digital data were tested scanning and digital microscopy. The benefits and limitations of these methods are discussed below.

Scanning

The thick sections were scanned using *Epson® WF 2540 scanner* using 1600–2400dp resolution. Examples of the scanned images are shown in Figure 3.13. The scanned images were of

Figure 3.11: Photographs of the processed vessels.

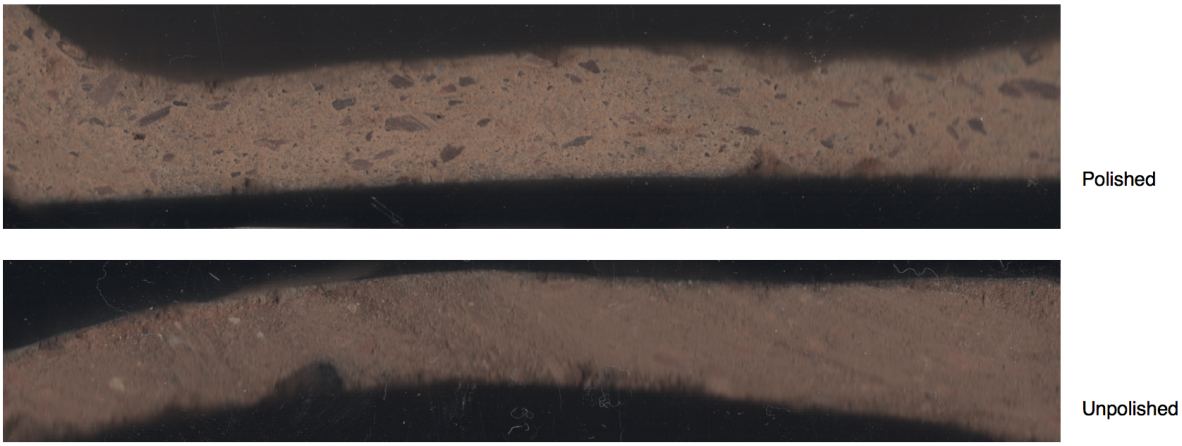


Figure 3.12: Examples of scanned ceramic thick sections.

high resolution and clarity, but due to their irregular shape they were challenging to digitally section into equally sized tiles to build a dataset of multiples examples.

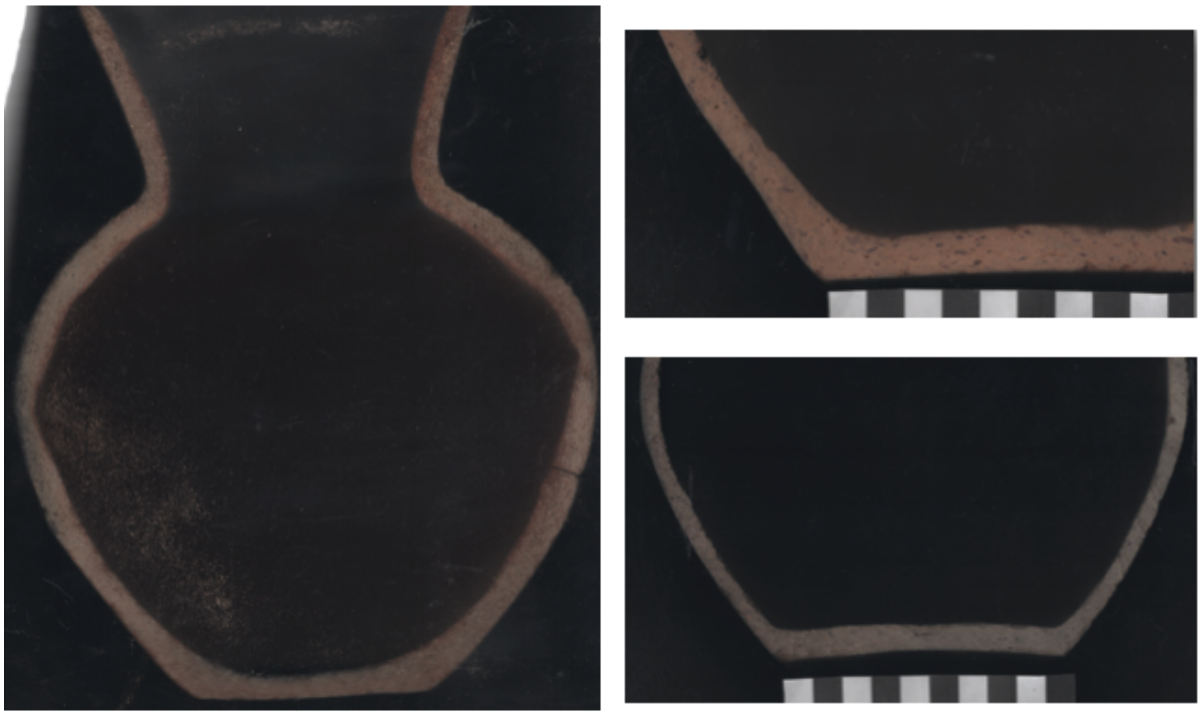


Figure 3.13: Examples of scanned thick sections before they were digitally cut into multiple tiles.

The large scans had to be digitally processed to generate multiple smaller tiles from each scan. First, it was attempted by manually selecting small parts of an image with a cursor and saving it as a separate image file. This method resulted in data inconsistencies, as there may be some overlap in areas being hand selected or parts of an image that could be easily missed. But more importantly, hand selecting and cutting out tiles in this way, is incredibly time consuming.

An alternative that worked well was to split a large scan into equal numbers of tiles and then

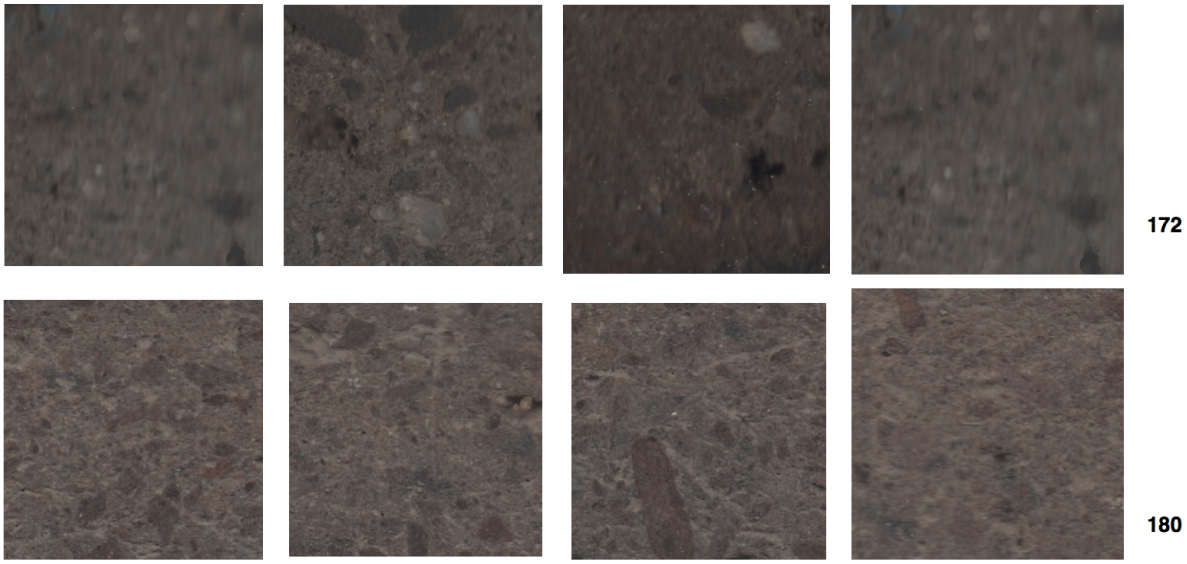
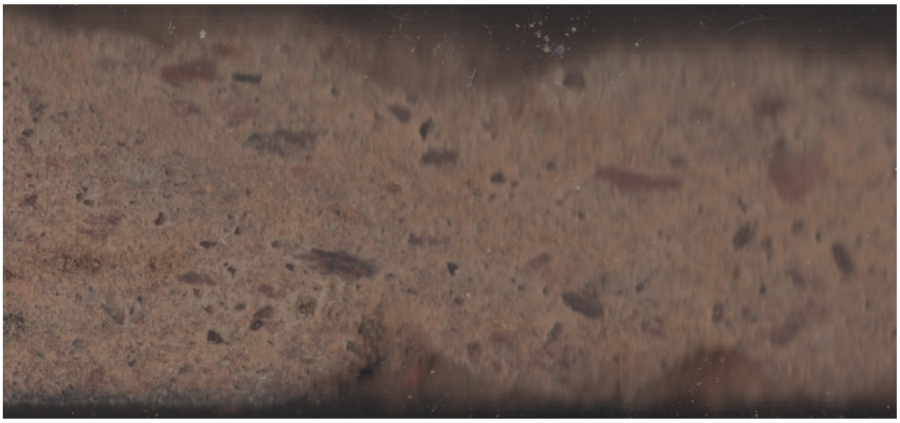


Figure 3.14: Examples of tiles produced from scanned thick sections.

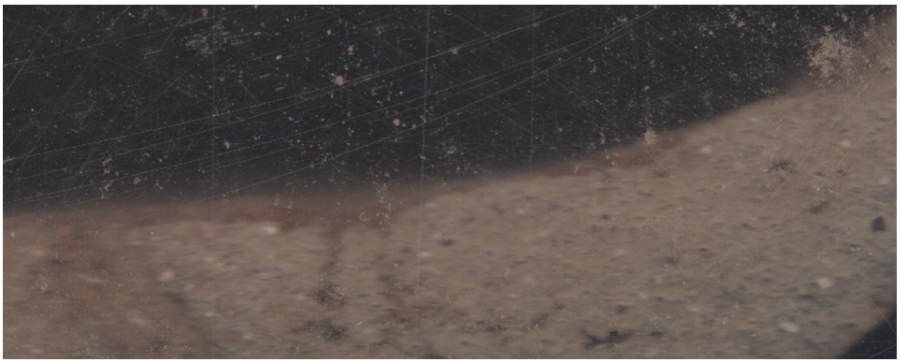
sort the resulting tiles by discarding tiles with a black background. This approach resolved the concern with overlap in an area being selected and was more time efficient as well. A *Python* script was used to split the original scans of various sizes into a maximum number possible of 320x240 tiles. Unfortunately, this was only attempted later in the project and experimenting with the obtained images was not included within the scope of this study.

This project identified several concerns associated with using the scanning method to obtain data for supervised machine learning analysis (Table 3.15). First, there was the presence of black shadows around the edges of the scans, causing variable range in the degree of blurriness between tiles. Second, some of the more pronounced marks from the saw were still visible post polishing. Especially in the case of the thicker marks left by the saw, a significantly more extensive polishing would be required to take off as much as 0.5 cm in most severe cases.



a.Blurry edges

b.Saw marks



c.Equipment scratches

Figure 3.15: Examples of scanned images with quality concerns.

The third concern with the scanning method is over the common equipment scratches due to use. Scratches on a scanner are inevitable to occur when the rough and hard surface of the ceramic material is placed on glass. These scratches are highly visible when high-resolution scans are taken and could impact the performance of an algorithm [8]. It is recommended to use the same equipment to take all images when possible. However, this may not always be feasible, for example, in cases where data being compared was acquired by different researchers working in different labs. A variation in density of scratches on different scanners could create noise for an algorithm. Being aware of that possible issue may help monitor for this problem when reviewing results.

Digital Microscopy

Due to the above challenges with scans, a decision was made to test an alternative method for collecting digital data. Next, the hand-held digital microscope *Dino-Lite 5MP Edge series AM7115MZTL* (Figure 3.16) was used to acquire images of the thick sections from the pots listed in Table 3.1.

Several challenges with using a hand-held digital microscope were identified. To acquire an image with this type of device a researcher must zoom in on one small surface area at a time while adjusting the magnification to focus the device. Once an image is taken, a researcher then needs to reposition the lens (by either moving the hand-held microscope or the thick section) onto an adjacent area while making slight adjustments to refocus the microscope before taking

[8] Equipment scratches can affect the performance of a classifier negatively, however, this requires further investigation to be conclusive.

Figure 3.16: Dino-Lite 5MP Edge series 11AM7115MZTL

Table 3.1: A list of pots photographed with a hand-held digital microscope.

Vessel Number	Region	Community	Type	Potter	Tiles
037	Nkandla	Magwaza	Umancishana	T. Magwaza	66
040	Nkandla	Magwaza	Ingazi	M. Magwaza	66
044	Nkandla	Magwaza	Umancishana	S. Magwaza	52
045	Nkandla	Magwaza	Umancishana	K. Magwaza	47
171	Msinga	Mabaso	Ukhamba	maMtungwa	64
172	Msinga	Mabaso	Ukhamba		99
177	Msinga	Mabaso	Ukhamba	Diza	88
180	Msinga	Mchuno	Ukhamba	MaMhuno	64
183	Msinga	Mchuno	Ukhamba	maMhuno	90

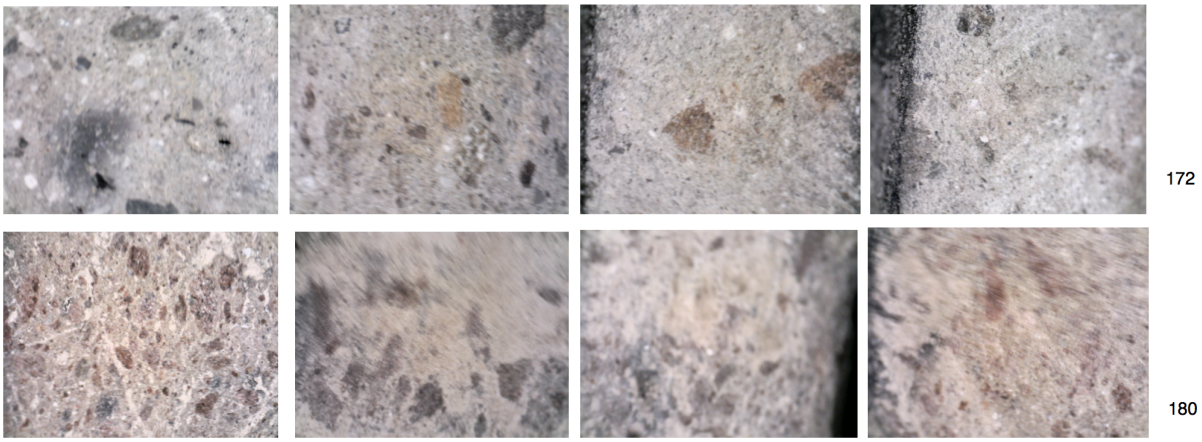


Figure 3.17: Examples of original 1280x960 images taken with *Dino-Lite 5MP Edge series AM7115MZTL* microscope.

the next image. This process is guaranteed to result in some inconsistencies. For example, repositioning the microscope precisely onto the following adjacent area is almost impossible. This can cause a partial overlap, where parts of two or more images taken to cover a part of the same surface area. Or, it can result in the opposite outcome, in which a researcher navigates further away from the last point when repositioning a lens, leaving a small gap, and excluding some area from being photographed.

Another major downside of a digital microscope is related to inconsistencies in the orientation of a hand-held microscope and an artifact during image acquisition. This only matters for particular classification problems, for which a consistent positioning of a microscope to acquire images from the same angle is essential. For example, orientation and position of inclusion and voids is a relevant variable to consider when classifying ceramics by shaping techniques.

One of the other challenging obstacles related to inconsistency is an inability to use the

same magnification settings while maintaining the same distance from which all images are photographed. Due to the variation in wall thickness of ceramics, or in size and coarseness of fabric inclusions, maintaining the same magnification and distance settings can result in many images being significantly more out of focus than others.

The high level of blurriness of an image, could be mistakenly interpreted by a trained classifier as relevant to classification. The primary goal is to ensure as much consistency as possible in both the image acquisition process, as well as in image quality. The decision was made to acquire data in a way that allows for flexibility in distance and magnification, but within a limited range. The images were obtained from a variable distance using a range of magnification. Table 3.2 below lists the specific details of magnification settings used for each vessel. However, even with allowing for flexibility in the distance and the magnification range to control for image quality, as is evident by the examples shown in Figure 3.17, there is still a significant degree of variability in the image quality.

Table 3.2: Magnification range used to acquire the images using *Dino-Lite 5MP Edge series AM7115MZTL* microscope.

Vessel Number	Magnification
37	x50 base / x70 walls
40	x70
44	x60
45	x65
171	x70
172	x50
177	x70
183	x70

And lastly, the black edges visible on some of the images are due to the variability in vessel's wall thickness. Changes in wall thickness within the same vessel as well as between vessels, resulted in this undesirable outcome. The thinner the vessel wall the more likely are the black edges to occur due to insufficient size of an area being photographed. The hand-held digital microscope works well for acquiring images of ceramic pots with a wall thickness of at least 0.5 cm. A portable microscope is not recommended when working with a collection of very small and fine pottery with thin walls.

Image Acquisition Choice

The main advantage of a hand-held digital microscope was that it eliminated the need for digital post-processing to split images into tiles. A handheld digital microscope collects multiple small images equal in size, making it a significantly more efficient approach compared to scanning. Given the above, the conclusion was made that a hand-held digital microscope was a more suitable method for acquiring images of the ceramic fabrics to create a dataset for machine learning analysis.

3.4.4 Datasets Design

A dataset can be defined as a collection of information that is organised in meaningful categories. This thesis looked at several classification schemes, including differences by paste, general shaping techniques, and minor differences between individual potters. Different dataset design strategies were required for each kind of classification problem. The datasets used in

this study are summarised here, with additional details available in the Appendix A.

Classification by Regional Differences in Paste

Paste differences in Zulu ceramics correlate to either an intended function of pottery or due to regional variation in where clays are sourced from and how those clays are processed and mixed. The majority of the vessels sampled are of the same functional type, but originating from different regions. Therefore the differences in paste being explored in this project are based on the latter. The datasets were broken down into categories intended to examine the differences in paste by region and by community.

The ability of *wndchrm* to distinguish minor regional differences in paste preparation process and clay sources was tested using two kinds of datasets. The first dataset is composed of two classes representing two regions Nkandla and Msinga (Table3.3), and the second dataset categorises examples by a community, including the Magwaza, Mchuno, and Mabaso communities (Table3.5). An alternative dataset was also created in which only the examples from the Mabaso and Mchuno communities were included (Table 3.4).

Table 3.3: Dataset composition by differences in paste by region: R001.

Class	Tiles Available	Vessel Number	Community
Nkandla	208	037 040 044 045	(Magwaza)
Msinga	405	171 172 177 180 183	(Mabaso, Mchuno)

Table 3.4: Dataset composition by differences in paste by community: C001.

Class	Tiles Available	Vessel Number	Region
Mabaso	251	171 172 177	(Msinga)
Mchuno	154	180 183	(Msinga)

Table 3.5: Dataset composition by differences in paste by community: C001.

Class	Tiles Available	Vessel Number	Region
Magwaza	208	037 040 044 045	(Nnkandla)
Mabaso	251	171 172 177	(Msinga)
Mchuno	154	180 183	(Msinga)

Classification by Shaping Technique

All the vessels were shaped using a combination of the coil and slab building techniques. Walls are representative of coiling, while bases are representative of slab building. In this datasets (Table 3.6), the images are categorised into two classes, as either coil or slab examples. No other information is provided to the *wndchrm* classifier. Table 3.6 lists the pots represented in the experimental datasets to classify fabrics by differences in the shaping techniques used.

Table 3.6: Dataset composition by differences in shaping technique: S001.

Class	Tiles Available	Vessel Number	Part
Coil	400	037 049 044 045 171 172 177 180 183 656	(Wall)
Slab	241	037 049 044 045 171 172 177 180 183 656	(Base)

Classification by Individual Potter

Several different datasets were designed to distinguish ceramics made by different potters. Some datasets focus on exploring the slight variations in shaping and hand movement patterns of individuals, while others were meant to account for paste differences.

More specifically three kinds of datasets were created. The datasets consisting of the sample taken from either the bases or the walls of a pot were designed to classify examples from different potters by shaping variability (Table 3.7, Table 3.8, and Table 3.9). The datasets that include samples from both parts of a vessel were designed to place more emphasis on paste differences (Table 3.10, Table 3.11, and Table 3.12). The samples used to create the training datasets to distinguish ceramics made by the particular potters are listed below.

Table 3.7: Dataset composition to identify minor difference in shaping by potter: IC001.

Class	Tiles Available	Vessel Number	Part	Region/Community
Potter A	80	171 172	Wall	(Msinga/Mabaso)
Potter B	67	180 183	Wall	(Msinga/Mchuno)

Table 3.8: Dataset composition to identify minor difference in shaping by potter: IC001.

Class	Tiles Available	Vessel Number	Part	Region/Community
Potter A	80	171 172	Wall	(Msinga/Mabaso)
Potter B	67	180 183	Wall	(Msinga/Mchuno)
Potter C	47	177	Wall	(Msinga/Mabaso)

Table 3.9: Dataset composition to identify minor difference in shaping by potter: IS001.

Class	Tiles Available	Vessel Number	Part	Region/Community
Potter A	50	171 172	Base	(Msinga/Mabaso)
Potter B	56	180 183	Base	(Msinga/Mchuno)

Table 3.10: Dataset composition to identify minor difference in paste by potter: IBW001.

Class	Tiles Available	Vessel Number	Part	Region/Community
Potter A	50	171 172	Base and Wall	(Msinga/Mabaso)
Potter B	56	180 183	Base and Wall	(Msinga/Mchuno)
Potter C	83	177	Base and Wall	(Msinga/Mabaso)
Potter D	41	044	Base and Wall	(Msinga/Mabaso)

Table 3.11: Dataset composition to identify minor difference in paste by potter: IBW001.

Class	Tiles Available	Vessel Number	Part	Region/Community
Potter A	80	171 172	Wall and Base	(Msinga/Mabaso)
Potter B	67	180 183	Wall and Base	(Msinga/Mchuno)

Table 3.12: Dataset composition to identify minor difference in paste by potter: BW001b_3.

Class	Tiles Available	Vessel Number	Part	Region/Community
Potter A	80	171 172	Wall and Base	(Msinga/Mabaso)
Potter B	67	180 183	Wall and Base	(Msinga/Mchuno)
Potter C	66	177	Wall and Base	(Msinga/Mabaso)

Classification by Individual Vessel

Several experiments were also conducted to identify samples from a particular vessel. In this case each class was represented by a single vessel. This classification scheme was particularly valuable as it featured the largest number of classes, which would help assess the performance of the *wndchrm's* classifier with as many as eight possible classes (Table 3.17) included in the same experiment. Dataset descriptions for classification by vessel are provided below.

Table 3.13: Dataset composition to identify samples from the same vessel: V001a.

Class	Tiles Available	Region/Community
040	49	(Nkandla/Magwaza)
044	41	(Nkandla/Magwaza)
171	56	(Msinga/Mabaso)
172	74	(Msinga/Mabaso)
177	83	(Msinga/Mabaso)

Table 3.14: Dataset composition to identify samples from the same vessel: V001b_5.

Class	Tiles Available	Region/Community
180	48	(Msinga/Mchuno)
183	75	(Msinga/Mchuno)
171	56	(Msinga/Mabaso)
172	74	(Msinga/Mabaso)
177	83	(Msinga/Mabaso)

Table 3.15: Dataset composition to identify samples from the same vessel: V001a_4.

Class	Tiles Available	Region/Community
180	48	(Msinga/Mchuno)
183	75	(Msinga/Mchuno)
171	56	(Msinga/Mabaso)
172	74	(Msinga/Mabaso)

Table 3.16: Dataset composition to identify samples from the same vessel: V001b_3.

Class	Tiles Available	Region/Community
040	49	(Nkandla/Magwaza)
044	41	(Nkandla/Magwaza)
177	83	(Msinga/Mabaso)

Table 3.17: Dataset composition to identify samples from the same vessel: V001b_8.

Class	Tiles Available	Region/Community
040	49	(Nkandla/Magwaza)
044	41	(Nkandla/Magwaza)
045	35	(Nkandla/Magwaza)
180	48	(Msinga/Mchuno)
183	75	(Msinga/Mchuno)
171	56	(Msinga/Mabaso)
172	74	(Msinga/Mabaso)
177	83	(Msinga/Mabaso)

3.4.5 Experiments Design

Further sub-variations of the datasets were created to explore best practices in dataset design and experiment with different training and testing parameters for the *wndchrm* model. The various options related to the number of examples used for training, image tiling, image resolution, and image quality were tested to understand the impact of some particular choices on the effectiveness of the classifier.

Image Quality

Three kinds of datasets were used to assess how image quality impacts *wndchrm's* performance. The first dataset consisted of images that were manually selected by the researcher as best in quality. The second dataset had the images with dark edges and excessive dark spaces excluded. The third dataset had the pictures which were deemed by the researcher as very blurry and where eliminated while keeping the images with dark edges. The goal of these experiments was to assess how variability in image quality impacts *wndchrm's* performance.

Accuracy and Size of Dataset

The number of training examples is one of the critical considerations that is likely to impact the accuracy. There is no one size fits all approach to know in advance how many training examples will be required for classification of ceramic fabrics. This thesis will experiment with different dataset sizes to assess how many training examples may be necessary.

The number of training examples required varies greatly depending on the quality of a dataset, the degree of variation between different classes, and the number of classes, as well as the kinds of features used. The more similar images are within a class and different between classes, the fewer examples should be required (Shamir et al., 2010). The accuracy is expected to increase incrementally as more training examples are added, eventually reaching a point where additional training examples no longer improve accuracy any further.

By incrementally reducing or increasing the numbers of examples used for training, one can observe whether there is a parallel incremental impact on attainable accuracy (Shamir et al., 2010). This can inform whether the lack of good results could be due to an inadequate dataset with too few training examples.

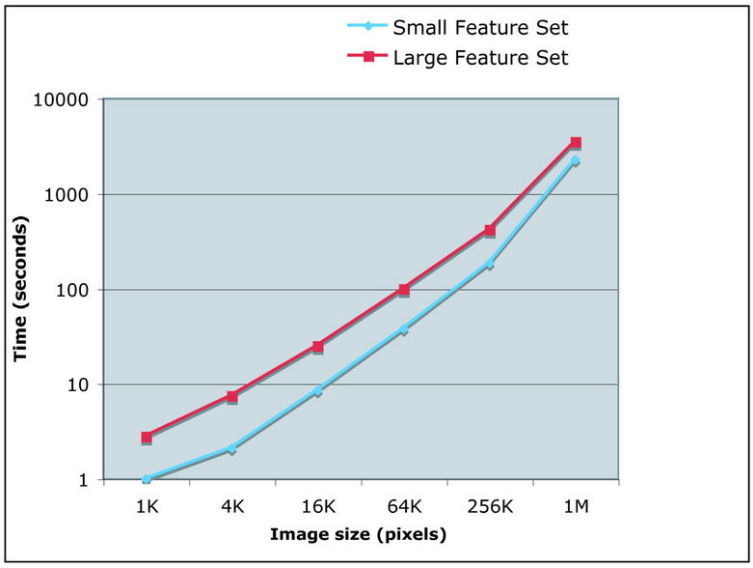
Tiling

Tiling can be used to split original images into multiple smaller images to increase sample size. However, there is a trade off between having a larger training set in exchange for using images representing a smaller proportion of a pot. For example, if the preferred orientation of inclusions is of interest, then generating smaller tiles might not show enough voids and inclusions to allow the assessment of their preferred orientation. Two kinds of datasets were used to learn the impact of tiling on classification outcomes. The first dataset consisted of images that were produced by splitting one original image into four equal sized tiles. The second dataset was built by dividing each original image into 15 tiles.

Feature Extraction

Wndchrm allows for two options when it comes to feature extraction. Researchers can choose to either extract a smaller set of 1025 features, or a larger set of 4059 features. To extract a large set of features a longer processing time is required. Figure 3.18 shows how the size of images and numbers of features being extracted effects the processing speed (Shamir et al., 2008). Extraction of the colour features requires an additional processing time as well. Colour may or may not have a significant impact on accuracy. Using the small set of features, the large set of features, the colour features, and the combination of the large set of features with the colour features was tested in multiple experiments using the same dataset, so that the results can be compared.

Figure 3.18: Expected increase in processing time when large vs. small sets of features are extracted by *wndchrm*. (Shamir et al., 2008)



.

Image Size

A notable decrease in processing speed is also linked to the image size, with reports of images over 512 x 512 pixels (Shamir et al., 2008) taking significantly longer to process. The larger the image being processed, the longer it takes. Three kinds of image resolutions were tested, including 1280x960, 640x480, and 320x240.

3.5 Understanding the Results

3.5.1 Results Output

Wndchrm provides several options for obtaining the results of the classification experiments. By default, results are displayed in a command line utility after *wndchrm test* or *wndchrm classify* command is completed. The output in a command line utility, however, is of limited formatting, making it harder to read and share. An alternative option is available in the form of an HTML report file summarising details of conducted tests. A report file includes details such as:

- Composition of the dataset, classes and examples represented in each class.
- Accuracy summary, as well as details of all splits when applicable.
- Further details on individual test images, indicating how each was classified, and similarity value of each image to different classes.
- Specific features used and their weight values, and rank.

- Class Probability Matrix Matrix.[9]
- Similarity Matrix.[10]
- Confusion Matrix.[11]
- Phylogeny Graph

3.5.2 Accuracy and Statistical Significance

Accuracy is defined as a percentage of unlabelled examples that were identified correctly out of all the tested items.*Wndchrm* allows for conducting tests on the same dataset in multiple random splits, which is particularly useful for cross-validating a classifier when dealing with a small dataset. With each split, a dataset is randomly divided into a different combination of training and testing examples. The average accuracy is calculated as an average of all splits. All the experiments in this project were based on ten splits.

The *confidence level* used is 95 percent, with a corresponding *significance level* of 0.05 percent. A *margin of error* (also known as the confidence interval) indicates +/- how many percentage points the results may differ if another similar experiment using a different set of examples was to be conducted. The Margin of Error (confidence interval) is calculated as follows:

[9] The class probability matrix is the average of marginal probabilities for the images in each class.
[10] The similarity matrix is the class probability matrix, where each row is normalized to make the class identity column equal to 1.0 The dis-similarity (i.e. 1.0 - similarity) between two classes can be interpreted as a "morphological distance There are two entries in the similarity matrix for each comparison: Class 1 classified as Class 2, and Class 2 classified as Class 1.
[11] Shows how many examples from Class 1 were missclassifed as Class 2, and vise versa.

$$Marginof Error = Zscore(1.96) * StandardErrorofMean(SEM)$$

Z-score for 95 percent confidence is 1.95996. The *Standard Error of the Mean* (SEM) measures how far the sample mean is likely to be from the true mean of a population.

Typically, when the accuracy reaches over 80%, the model should be considered as being able to discriminate between classes, even when as few as two classes are involved. However, there is no particular threshold for assessing whether the accuracy should be considered as high. As a general rule, if a machine learning model is able to achieve better results than a human expert or other existing methods, it should be viewed as useful. Assessment of how human experts compare in their ability to identify ceramic fabrics was not included as part of the scope of this project, but can be a worthwhile study in the future. The requirement for accuracy varies depending on the nature of the problems and questions that are being investigated. Some types of interpretations may require a higher degree of accuracy than others.

Chapter 4

Results

This chapter summarises the results of the conducted experiments and notes how different dataset design and model training parameters can impact the accuracy of classification. Appendix A summarises the details of all the experiments conducted as part of this study. The analysis in this thesis project involved a small number of vessels with a sample size significantly increased by generating multiple images from each vessel. It is recommended that future studies use more pots to verify the validity of the presented findings further.

4.1 Experimental Results

Regions and Communities

Wndchrm was able to classify the ceramic fabrics by community with an accuracy of 94% (Table A.7, (Table A.13, (Table A.14). The experiments were based on two datasets. The first consisted of two communities Mabaso and Mchuno (**??**). The second included the ceramic examples from Mabaso and Mchuno, as well as Magwaza. The dataset consisting of three classes required a higher number of the training examples, but the experiments based on both datasets

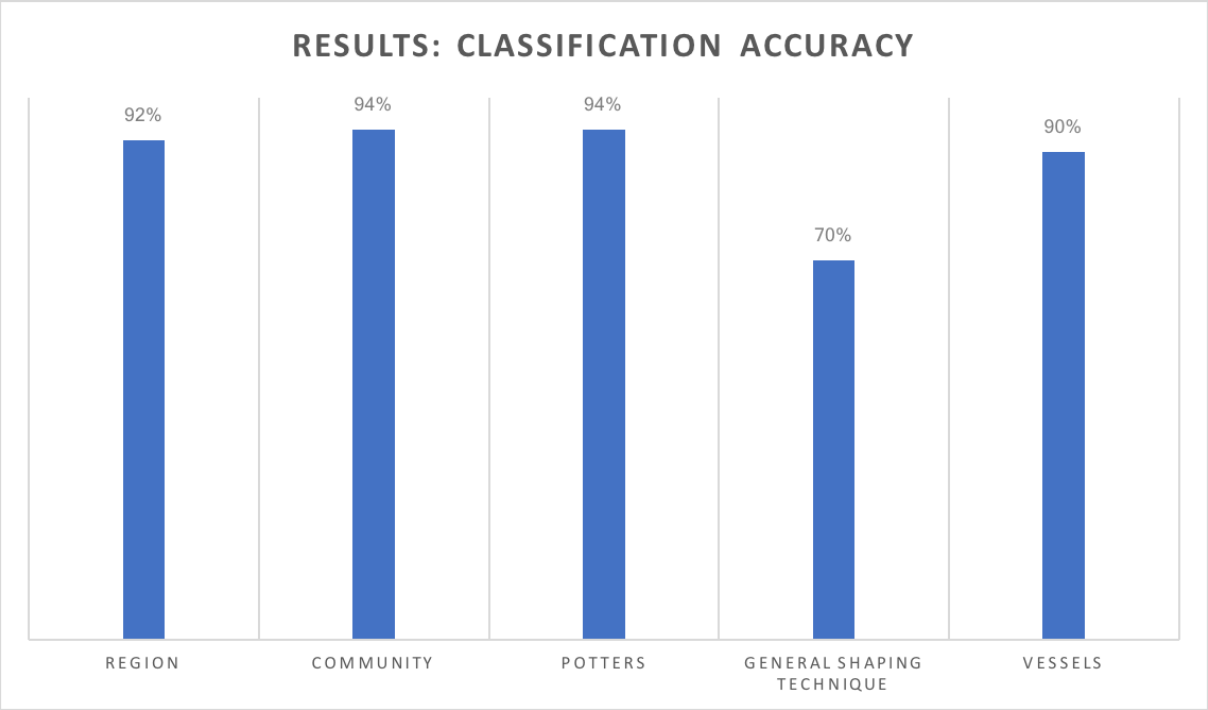


Figure 4.1: Attainable accuracy by classification problem.

showed the accuracy of 94% (Table A.14). Fifty training images per class were required for the three class experiments compared to the two class experiments which needed 30 training examples per class to reach the same level of accuracy.

Another two class dataset was created to classify the images by region, as either Nkandla or Msinga. The experiments using this dataset were also successful (Table A.6). *Wndchrm* was able to predict the region of origin for the unlabelled examples with an accuracy of 92%. Classification experiments by region required at least 60 training examples to achieve the same level of accuracy.

Vessels

Several different datasets were used to identify the ceramic samples taken from a particular vessel. The lowest performance was observed when the dataset consisting of the mixed vessels from Msinga and Nkandla was used. The highest accuracy achieved in this case was 77%. The dataset used for this particular experiment consisted of 8 classes, with 30 training examples represented per class (Table A.20). Given the observed correlation between the accuracy gains and the increase in the number of training examples, it is likely that the accuracy will continue to improve as more training examples are added.

Other datasets consisted of vessels taken from the single region (Msinga) and produced significantly better results, reaching the accuracy of 85% (Table A.18). This result is also likely to improve with the increase in the number of training examples. The highest number of training examples tested for these experiments was 40 examples per class, with five and four classes represented respectively in each dataset.

Shaping Techniques

Wndchrm was not able to distinguish as successfully between the slab building and coiling shaping techniques. Classification by the shaping method produced the highest accuracy of 65% to 70% (Table A.21). To reach this level of accuracy 100 training examples per class were required. No significant improvement in the accuracy was observed past 100 training examples, suggesting that increasing the sample size will not result in further performance improvement.

Tiling of the original images to generate a larger training set resulted in decreased accuracy, bringing it down to 60% (Table A.22). This suggests that using pictures representative of a greater surface area of a pot could potentially increase accuracy. Therefore, it is recommended to prioritise using fewer images representing a more substantial part of a pot, over a higher number of images that represent a smaller portion of a pot.

Individual Potters

The experiments designed to classify the fabrics by potters were initially intended to consider the samples taken from either the bases or the walls in separate experiments, using different datasets. This split was meant to identify the variability between the individual potters based on the minor differences in shaping. *Wndchrm* was able to classify the new examples with 90% accuracy in both cases (Table A.23 and Table A.29). These results were unexpected given *wndchrm's* low performance in categorising the ceramic fabrics by general shaping techniques.

To test whether *wndchrm's* ability to successfully distinguish between potters was based on the variability in the paste preparation processes rather than due to the potters' unique hand movement patterns during the shaping of the vessels, an additional dataset was created combining the samples taken from the walls and the bases together. The experiments which included both the walls and the bases of the pots as part of the same training set also achieved the accuracy of 90% (Table A.32, Table A.33, and Table A.34). This suggests that the *wndchrm* classifier is effective in identifying minor differences in paste preparation particular to an individual. One of the potters (Potter A) represented in the experiments used a stone in a rolling

motion to grind clay, while another potter (Potter B) used a beating technique, which is an example of the differences in the paste preparation practices between the two potters.

4.2 Highest Ranked Features

The majority of the experiments consistently placed increased emphasis on the Heralick Texture group of features. Other groups of features were deemed as relevant only in particular cases. For example, the Fractal features played an important role in classifying the ceramic fabrics by the individual potters or vessels. The experiments designed to distinguish between the potters including the examples taken from walls of the vessels specifically placed higher weight on the Tamura texture features. Gini Coefficient (Hue) was distinctively more relevant to categorise the fabrics by a community.

An observation worth noting relates to the Pixel Intensity features. This group of features was ranked high only in the experiments which included the samples from the Mabasa Potter C. *Wndchrm* was able to identify the examples from Potter C with 100% accuracy in multiple trials, even in the case of using 10 training images and 56 test images from this potter. This finding is very unusual, and upon further investigation, *wndchrm* assessed the Potter C as being very dissimilar to any of the other potters tested from the same region or community. All the Zulu potters used very similar raw materials and methods to mix the paste. The only notable difference in regards to the Potter C is that she is a novice potter. It is plausible that *wndchrm* was able to successfully distinguish between the novice and experienced potters based on their

abilities to process clay and mix paste.

The experiment designed to classify fabrics by region was the most diverse in the types of features that were given increased emphasis. Some of the examples of the features that were ranked high for regional classification include Gini Coefficient (Hue), Zernike Coefficient (Hue), Heralick Textures (Wavelet Fourier), Comb Moments (Hue), Multiscale Histogram, Tamura Textures, and Pixel Intensities. The highest weighted features for different classification problems are listed in Appendix B (Table B.1).

While task specific algorithms can be built to account for particular image attributes of interest to the human expert, *wndchrm* is a general purpose algorithm meant to be applied to a broad variety of images. As a result a large number of features are extracted, many of which are abstract in nature or numerical representations not directly translatable to obvious graphical attributes that can be linked back to the original image. The different groups of features described fall into one of the four categories. Polynomial decompositions are the closest approximation of an actual image. The other three are pixel statistics, textures, and high contract features consisting of statistics about edges and objects such as spacial distribution, size, or shape. Many features derive from image transforms which further distort the connection to the original graphical content.

4.3 Training Parameters

The best accuracy was achieved consistently for all of the experiments when the colour features and the large set of features were included. The use of the colour features was an especially significant factor for the experiment R001a that classified data by region. The inclusion of the large set of features as well as the colour features resulted in improvements in accuracy by an average of 10% for most of the experiments. The extraction of the additional features took three times longer, however given the correlated gain in accuracy, it is worth the extra time.

4.4 Testing Parameters

Multiple experiments were conducted to determine the optimal proportion of features to be used by the classifier, ranging from 2% to 40% (Table A.6, Table A.7, Table A.6, Table A.13, Table A.14, Table A.15, Table A.17, Table A.21, Table A.23, Table A.29, Table A.33). No notable correlation between the percentage of selected features and attainable accuracy was observed. Using the nearest neighbour approach instead of the nearest distance did not impact accuracy either. However, the closest neighbour (- w) variation of *wndchrm* took significantly longer to process, especially on the larger datasets, and therefore it is not recommended.

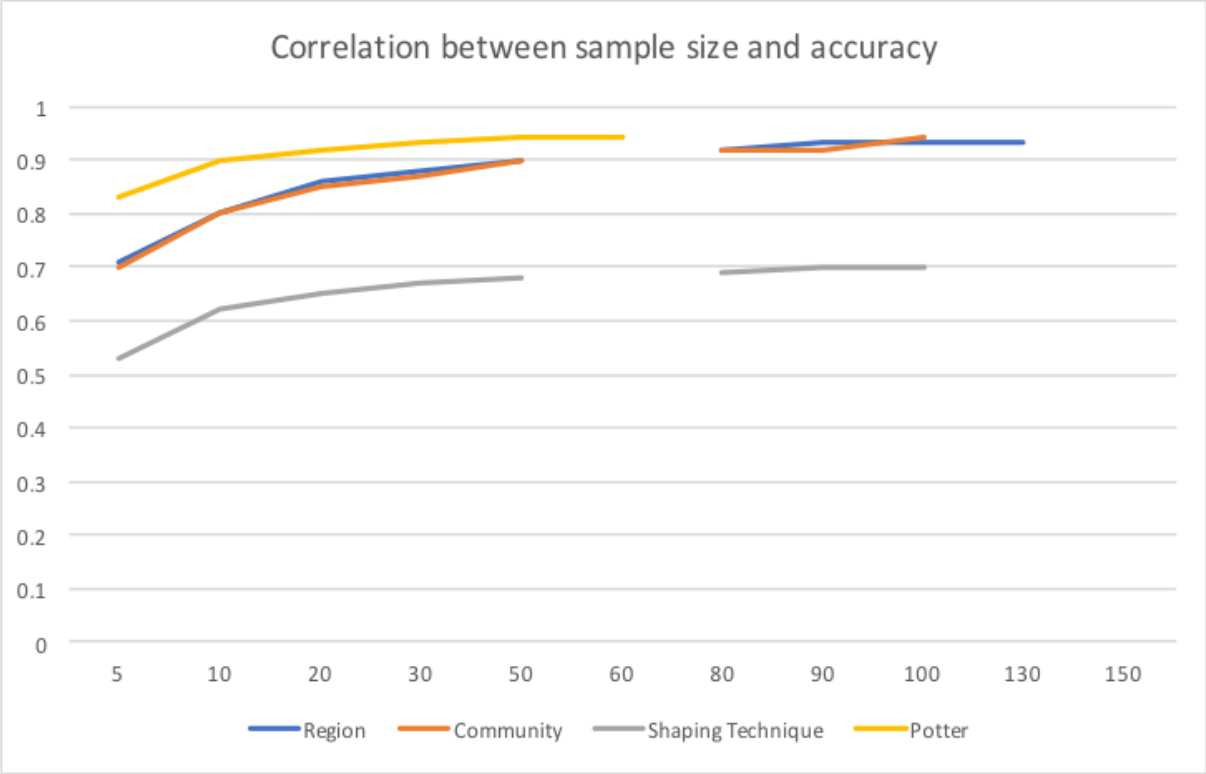


Figure 4.2: Training set size and accuracy

Table 4.1: The number of training examples per class and attainable accuracy.

Training Examples (per class)	Region	Community	Shaping Technique	Potter
5	71%	70%	53%	83%
10	80%	80%	62%	90%
20	86%	85%	65%	92%
30	88%	87%	67%	93%
50	90%	90%	68%	**94%**
60				94%
80	92%	92%	69%	
90	**93%**	92%	**70%**	
100	93%	**94%**	70%	
130	93%			
150			70%	

4.5 Dataset Design

Numbers of Training Examples Required

To distinguish the samples by individual potters, the training set of 20 examples per class was sufficient to achieve an accuracy of 90%. A slight gradual increase in accuracy was observed when more training examples were added. After 60 training examples per class, there was no more notable improvement and using more than 60 training images is not likely to further improve these results.

In contrast, to identify fabrics from the same vessel, a minimum of 30 training images per class were required. The accuracy is expected to increase with a larger sample size in this case, and ideally, more than 30 training images should be used.

To classify the ceramic fabrics by community, a set of 50 training images per class was required to obtain the average accuracy of 80%. The increase in the sample size showed improvement in the results reaching as much as 94% accuracy when 100 training images per class were used. No additional improvement is expected past 100 training examples per class when classifying the Zulu fabrics by community.

Wndchrm demonstrated an ability to classify the samples by region with 80% accuracy, using 15 training images from each region. Gradually increasing the number of training images resulted in the linear increase in accuracy reaching 90% when 100 training examples were used.

Wndchrm was not very effective in distinguishing the images of ceramic fabrics by general

shaping technique. The highest accuracy of 70% was obtained with 150 training examples per class. This result is not expected to improve if the number of training examples was to increase.

Image Quality

Three kinds of datasets were tested. The first dataset was manually curated for image quality (Table A.7); images deemed to have significant blurriness or excessive dark edges were removed. The second dataset had blurry pictures eliminated, while those with the dark edges were kept (Table A.9). The third dataset had the pictures with the dark edges removed, but the blurry images were kept (Table A.10). No significant impact on the accuracy was observed and all three kinds of datasets performed equally well. Manually eliminating poor quality images of ceramic fabrics obtained using a digital microscope does not appear to be necessary.

Image Acquisition

Digital microscopy is the recommended method for obtaining images to classify ceramic fabrics by paste recipes. *Wndchrm* was able to successfully classify the images of ceramic fabrics by paste despite the concerns with inconsistent distance and magnification used to obtain the pictures with the handheld digital microscope. This tool, however, is not suitable for acquiring images of ceramic fabrics for identification of different shaping techniques.

Image Resolution

The study also investigated how image resolution impacts *wndchrm's* ability to classify ceramic fabrics. Some experiments were conducted using the datasets consisting of the images with 640 x 480, and 320 x 240 image resolutions (Table A.6, Table A.8, Table A.6, Table A.23 and Table A.25). Both performed equally well. Using 320 x 240 images resolution is recommended because lower resolution images require significantly less time to extract image features from.

Image Tiling

Tiling the original images to increase the sample size had a negative impact on the results. The datasets which consisted of images that were split into four equal squares from the original produced comparable, or in some cases worse results, compared to the original dataset Table A.11,. Some other experiments used the images split into 15 tiles from the original image (Table A.12. These datasets included hundreds of training examples per class. Despite a large number of training examples used, this approach had a negative impact on the results.

Tiling of images to increase sample size is not recommended in the case of ceramic fabrics classification. Higher accuracy is achievable with fewer examples of larger samples. Tiled datasets with a large number of training examples are also very expensive in processing time required. Extracting the large set of features from the dataset R001c [1] for example, took three full days compared to the dataset R001a consisting of the original images, which took five

[1] R001c is the original R001a tiled by 15

hours.

Research Time

Table 4.2: Data processing time.

Task	Research Time	Details
Physical Sample Processing	2 hrs per vessel	Cutting, polishing
Image Acquisition	30 minutes per vessel	Dino Lite Microscope
Image Tiling	not required	Microscopy Images
Image Acquisition	1 - 3 hours per vessel	Scanning
	1 - 3 hours per vessel	
Image Tiling	2 - 8 hours per vessel	Scanned Images

Research time required for the analysis greatly varies depending on the methods and tools used to collect and process data. Table 4.2 provides a general time estimate required for collecting data using different methods tested in this thesis. The most substantial time required however is in processing of the images to extract features. The time required varies greatly depending on what computer equipment is used. Extracting a large set of colour features from 276 images with a resolution of 640 by 480 for example took just over 4 days using a standard laptop. In contrast, extracting the same features using the latest MacBook Pro with 3.3 GHz Intel Core i7 processor took 6 hours. Extracting features from images that were 1280 by 960 in resolution close to doubled the feature extraction time, while reducing the size to 320 by 240 resolution cut that time in half. Further tiling of original image tiles obtained with a digital microscope produced large datasets which were highly computationally expensive to process.

Even when using a more powerful computer, equipment datasets that contained either four times or fifteen times more examples took a very long time to process. Tiled datasets took over a week to process and produced results with lower accuracy than the original images obtained with the digital microscope despite the larger number of training images being made available.

Chapter 5

Conclusion and Future Directions

Ceramic artifacts are one of the most abundant kinds of material found at archaeological sites. Pottery is commonly used by archaeologists to make interpretations about many aspects of life in the past. By analysing and comparing pottery remains within and between archaeological sites, researchers make interpretations about the economy and trade, the organisation of technology, the division of labor, and the craft specialisation (Balfet & Matson, 1965; Blackman, Stein, & Vandiver, 1993; Hangstrum, 1985; Lindahl & Pikirayi, 2010; Rye, 1981; Sinopoli, 1991).

Stylistic and morphological aspects of pottery can be relayed to other researchers through descriptions and photographs of different styles, or measurements and drawing of morphological pottery types. However, when it comes to exchanging information about ceramic fabrics, researchers face a significant challenge because it can not be described in the same way as morphological details. Supervised machine learning is an excellent addition to be used alongside of other research methods used in ceramic analysis. *Wndchrm* specifically can add significant value in this area of research.

As demonstrated in this thesis *wndchrm* can accurately distinguish ceramic fabrics by dif-

ferences in paste. Despite the fact that the general practices in clay processing and tempering are relatively homogeneous across Msinga and Nkandla regions, *wndchrm* was able to successfully pick up on even very minor variations, specific to a particular community and to individual potters. The ability of *wndchrm* to distinguish ceramic fabrics with 90% accuracy, given such minor variability in paste preparation, is what makes *wndchrm* particularly valuable to this field.

Next Steps

Curating training datasets, which can be shared for cross-comparison, is the next step for those archaeologists wishing to adopt this approach. There are a number of improvements to the field that can be made by adopting machine learning. It will bring consistency in comparing data between different collections. It will also reduce costs and increase speed of analysis. Machine learning can potentially reduce the frequency of expensive scientific analysis often required to interpret ceramic data. For example, archaeologists use Neutron Activation Analysis and X-Ray Diffraction to learn chemical signatures of clays or Polarising Microscopy to study optical mineralogy. A limited numbers of samples from a collection are typically tested using these methods as these analysis are time consuming and costly. Previously analysed ceramics with such methods can be used to build training datasets which can be used to potentially train the algorithm to classify the unknown data.

To begin disseminating training examples, one can simply create and share training sets from their ceramic collections. To build a training dataset, one needs to select examples from their assemblage, documenting the details about each training class, listing everything that is

already known about the fabrics, such as their context and any findings from previous analysis.

To classify new samples using a pre-existing training dataset from another collection, the unlabeled samples should be placed together in a separate *test* folder. *Wndchrm* can then be used to classify the new data in that folder, using the relevant features informed by the training dataset. *Wndchrm* quantifies similarity values which analyse the results and search for the most similar fabrics using the different training sets available.

Creating a reference collection of training datasets is the first step towards adopting supervised machine learning for classification of ceramic fabrics. Training datasets are not restricted to *wndchrm* and they can be used with other algorithms and approaches as well. Therefore, curating a library of training datasets is a worthwhile investment for the future of research in this field. It is important to ensure some degree of consistency exists between the datasets for them to be compared effectively. The next section provides guidelines for creating a dataset that is optimised for accuracy as well as for speed of analysis.

Dataset Recommendations

The hand-held digital microscope is a practical and time-efficient tool recommended for acquiring images of ceramic fabrics to build a dataset to classify ceramics based on paste differences. The handheld digital microscope also works with smaller sherds, which is the most common kind of ceramic material found during an archaeological excavation. Table 5.1 summarizes recommendations for building an effective dataset and Table 5.2 lists model training parameters particular to using *wndchrm* for classification of ceramic fabrics by differences in paste.

Table 5.1: List of recommendations for building a dataset to classify ceramic fabrics by paste differences.

Magnification Range:	x40 - x75
Image Quality:	include all images acquired
Image Resolution:	320x240
Tiling:	not recommended
Min. Training Examples:	10
Recommended Training Examples:	40
Max. Training Examples:	100

Table 5.2: List of recommendations for training *Wndchrm* model to classify ceramic fabrics by paste differences.

Feature Extraction:	large set of features, colour features
Method:	weighted distance
Feature Selection Ratio:	variable, no significant impact was observed

Further Research on Differences in Shaping Technique

The classification experiments intended to distinguish between the shaping techniques did not produce accurate enough results to be useful. Assuming that the orientation of inclusions is essential for identifying shaping techniques, the method used to obtain images in this project is not a recommended approach, mainly due to the inconsistencies in positioning discussed in detail in Chapter 3. For any other studies wishing to apply supervised machine learning to classify ceramics by differences in shaping techniques, it is advisable to use images representing a larger section of a vessel. Scanning thick sections of complete pots is recommended. Scans are likely to contain significant dark regions. However, there does not appear to be any significant impact on accuracy when examples with dark edges are included in datasets. Scanning

images at the lowest resolution possible is recommended to reduce the visibility of equipment scratches. Furthermore, high image resolution has no notable impact on the attainable accuracy, but requires longer processing time.

X-ray is another method for image acquisition that can be experimented with when complete or nearly complete pots are available. The x-ray images will also allow for more consistency in artifact positioning, as well as a full view of the vessel. X-ray images can also allow for analysis of complete pots without having to cut them, which is critical for studying any rare material derived in an archaeological context. X-ray, however, is time-consuming and not easily accessible for all researchers. Methods that are time efficient and easily accessible are likely to have the most impact and should be explored first.

Other Approaches

The *wndchrm* approach tested is an example of supervised learning in which there are examples with known class memberships. Christian Horr discussed how this translated to archaeological applications specifically, where we want with interpreting a collection of unknown objects. He argues that, in most cases there is some expert insight that exists, providing a starting basis by identifying a few obvious examples that can be identified (Hörr et al., 2014, 1). Horr's work on classifying ceramics by morphological types provides an example of applying a semi-supervised approach to archaeological problem. Depending on the nature of the question and data at hand, one can choose to apply either a supervised, unsupervised, or semi-supervised approach.

Appendix A

Experiments Summary

The results of the conducted experiments are listed below. Each result value listed respresents a separate experiment. The analysis are based on the 95 percent confidence level, and the the margin of error is also provided.

Due to the number of considerations and experimentation in regards to the image sizes, quantity, quality in the dataset design, as well as multiple experiments using different training and testing parameters, a labelling system to document the datasets and experiments was necessary. The labelling system used is summarised below.

Table A.1: Dataset and experiments: documentation and labelling

Code	Description
R	region
C	community
S	shaping
IS	individual, slab
IC	individual, wall
V	vessel

Table A.2: Dataset design and labelling details: image quality

Code	Tile Selection
001	best quality
003	best quality + dark edges
004	best quality + blurry

Table A.3: Dataset design and labelling details: image resolution

Code	Resolution
a	640x480
b	320x240
o	1280x960

Table A.4: Dataset design and labelling details: tiling

Code	Original	Tile Resolution	Numbers of Tiles per Original
c	1280x960	320x240	15
d	1280x960	320x240	4
e	640x480	320x240	4

Table A.5: Model testing and training parameters labelling.

Code	Description
default	small set of features (no colour)
L	large set of features
C	colour was used
LC	both colour and large set of features was used
i	number of training images per class
f	ratio of features considered out of all computed
w	weighted neigbour was used

Table A.6: Experiments: Region (R001a)

Train: 100 (per class) Test: 51 (per class)				
	default	**-L**	**-C**	**-L -C**
f.40 -	82.5 +/- 2.3%	86.0 +/- 2.1%	92.7 +/- 1.6%	91.7 +/- 1.7%
f.35 -	84.0 +/- 2.2%	83.7 +/- 2.3%	92.1 +/- 1.7%	93.8 +/- 1.5%
f.30 -	85.5 +/- 2.2%	87.1 +/- 2.1%	93.8 +/- 1.5%	92.4 +/- 1.6%
f.25 -	85.0 +/- 2.2%	85.9 +/- 2.1%	92.3 +/- 1.6%	91.6 +/- 1.7%
f.20 -	84.9 +/- 2.2%	85.1 +/- 2.2%	93.4 +/- 1.5%	91.5 +/- 1.7%
f.15 -	84.5 +/- 2.2%	83.2 +/- 2.3%	93.3 +/- 1.5%	93.4 +/- 1.5%
f.10 -	82.0 +/- 2.4%	82.9 +/- 2.3%	92.5 +/- 1.6%	90.6 +/- 1.8%
f.05-	85.6 +/- 2.2%	83.2 +/- 2.3%	90.9 +/- 1.8%	92.1 +/- 1.7%
f.02-	81.4 +/- 2.4%	82.9 +/- 2.3%	91.7% +/- 1.7%	93.2 +/- 1.5%
Train: 130 (per class) Test 21 (per class)				
f.40 -	85.2 +/- 3.4%	89.3 +/- 3.0%	91.7 +/- 2.6%	94.0 +/- 2.3%
f.30 -	83.3 +/- 3.6%	85.7 +/- 3.3%	93.6 +/- 2.3%	92.1 +/- 2.6%
f.15 -	88.1 +/- 3.1%	82.1 +/- 3.7%	91.2 +/- 2.7%	92.1 +/- 2.6%
Train: 90 (per class) Test 61 (per class)				
f.40 -	82.3 +/- 2.1%	85.2 +/- 2.0%	91.4 +/- 1.6%	93.1 +/- 1.4%
f.30 -	83.9 +/- 2.1%	83.6 +/- 2.1%	91.6 +/- 1.6%	92.8 +/- 1.5%
f.15 -	82.6 +/- 2.1%	83.5 +/- 2.1%	93.1 +/- 1.4%	91.8 +/- 1.5%
Train: 80 (per class) Test (71 per class)				
f.40 -	84.4 +/- 1.9%	84.6 +/- 1.9%	91.9 +/- 1.4%	92.0 +/- 1.4%
f.30 -	82.6 +/- 2.0%	83.9 +/- 1.9%	92.6 +/- 1.4%	92.0 +/- 1.4%
f.15 -	84.1 +/- 1.9%	84.4 +/- 1.9%	92.5 +/- 1.4%	93.2 +/- 1.3%
Train: (50 per class) Test: (101 per class)				
f.40 -	82.6 +/- 1.7%	83.8 +/- 1.6%	90.1 +/- 1.3%	90.7 +/- 1.3%
f.30 -	82.3 +/- 1.7%	83.2 +/- 1.6%	92.4 +/- 1.2%	90.8 +/- 1.3%
f.15 -	81.0 +/- 1.7% (	83.3 +/- 1.6%	91.8 +/- 1.2%	90.9 +/- 1.3%
Train: 30 (per class) Test: 121 (per class)				
f.40 -	80.5 +/- 1.6%	79.2 +/- 1.6%	88.6 +/- 1.3%	87.9 +/- 1.3%
f.30 -	79.6 +/- 1.6%	81.3 +/- 1.6%	89.6 +/- 1.2%	89.2 +/- 1.2%
f.15 -	78.3 +/- 1.6%	81.5 +/- 1.5%	89.0 +/- 1.2%	89.0 +/- 1.2%
Train: 20 (per class) Test: (131 per class)				
f.40 -	80.6 +/- 1.5%	80.1 +/- 1.5%	87.3 +/- 1.3%	86.4 +/- 1.3%
f.30 -	76.8 +/- 1.6%	76.9 +/- 1.6%	85.6 +/- 1.3%	86.8 +/- 1.3%
f.15 -	77.4 +/- 1.6%	77.6 +/- 1.6%	86.9 +/- 1.3%	87.7 +/- 1.3%
Train: 10 (per class) Test: 141 (per class)				
f.40 -	74.9 +/- 1.6%	75.4 +/- 1.6%	78.8 +/- 1.5%	80.0 +/- 1.5%
f.30 -	73.1 +/- 1.6%	73.2 +/- 1.6%	84.5 +/- 1.3%	82.4 +/- 1.4%
f.15 -	75.0 +/- 1.6%	72.5 +/- 1.6%	80.2 +/- 1.5%	78.8 +/- 1.5%
Train: 5 (per class) Test: 131 (per class)				
f.40 -	69.7 +/- 1.7%	69.3 +/- 1.7%	72.5 +/- 1.6%	71.9 +/- 1.6%
f.30 -	70.2 +/- 1.7%	66.5 +/- 1.7%	70.3 +/- 1.7%	72.2 +/- 1.6%
f.15 -	65.8 +/- 1.7%	66.3 +/- 1.7%	65.6 +/- 1.7%	72.0 +/- 1.6%

Tiling: Original
Resolution: 640x480
Selection: Best Quality

Classes:
Nkandla
Msinga

Table A.7: Experiments: Region (R001a)

Train: 100 (per class) Test: 51 (per class)				
	-W default	**-W -L**	**-W -C**	**-W -L -C**
f.40 -	79.1 +/- 2.5%	81.1 +/- 2.4%	92.7 +/- 1.6%	92.9 +/- 1.6%
f.35 -	80.2 +/- 2.4%	80.8 +/- 2.4%	93.9 +/- 1.5%	92.8 +/- 1.6%
f.30 -	79.6 +/- 2.5%	82.0 +/- 2.4%	93.7 +/- 1.5%	93.5 +/- 1.5%
f.25 -	79.9 +/- 2.5%	81.2 +/- 2.4%	93.9 +/- 1.5%	93.5 +/- 1.5%
f.20 -	79.6 +/- 2.5%	79.7 +/- 2.5%	94.2 +/- 1.4%	91.4 +/- 1.7%
f.15 -	80.8 +/- 2.4%	80.7 +/- 2.4%	92.7 +/- 1.6%	92.6 +/- 1.6%
f.10 -	81.1 +/- 2.4%	80.3 +/- 2.4%	92.5 +/- 1.6%	91.2 +/- 1.7%
f.05-	79.1 +/- 2.5%	80.2 +/- 2.4%	91.8 +/- 1.7%	90.8 +/- 1.8%
f.02-	77.6 +/- 2.6%	80.1 +/- 2.5%	89.5 +/- 1.9%	92.8 +/- 1.6%
Train: 130 (per class) Test: 21 (per class)				
f.40 -	86.0 +/- 3.3%	83.3 +/- 3.6%	93.3 +/- 2.4%	92.4 +/- 2.5%
f.30 -	85.7 +/- 3.3%	82.6 +/- 3.6%	93.1 +/- 2.4%	91.4 +/- 2.7%
f.15 -	85.5 +/- 3.4%	83.6 +/- 3.5%	91.9 +/- 2.6%	94.8 +/- 2.1%
Train: 90 (per class) Test: 61 (per class)				
f.40 -	85.3 +/- 2.0%	79.3 +/- 2.3%	93.4 +/- 1.4%	91.8 +/- 1.5%
f.30 -	83.9 +/- 2.1%	79.3 +/- 2.3%	93.1 +/- 1.4%	92.9 +/- 1.4%
f.15 -	82.5 +/- 2.1%	81.0 +/- 2.2%	92.5 +/- 1.5%	91.6 +/- 1.6%
Train: 80 (per class) Test: 71 (per class)				
f.40 -	83.5 +/- 1.9%	82.0 +/- 2.0%	92.6 +/- 1.4%	91.8 +/- 1.4%
f.30 -	82.4 +/- 2.0%	81.6 +/- 2.0%	93.7 +/- 1.3%	91.5 +/- 1.5%
f.15 -	83.9 +/- 1.9%	79.9 +/- 2.1%	93.0 +/- 1.3%	92.1 +/- 1.4%
Train: 50 (per class) Test: 101 (per class)				
f.40 -	83.0 +/- 1.6%	78.9 +/- 1.8%	91.5 +/- 1.2%	91.5 +/- 1.2%
f.30 -	81.3 +/- 1.7%	76.8 +/- 1.8%	90.2 +/- 1.3%	90.3 +/- 1.3%
f.15 -	81.0 +/- 1.7%	79.0 +/- 1.8%	90.6 +/- 1.3%	90.9 +/- 1.3%
Train: 30 (per class) Test: 121 (per class)				
f.40 -	79.3 +/- 1.6%	77.5 +/- 1.7%	89.8 +/- 1.2%	87.9 +/- 1.3%
f.30 -	79.4 +/- 1.6%	76.2 +/- 1.7%	90.0 +/- 1.2%	88.8 +/- 1.3%
f.15 -	78.6 +/- 1.6%	78.4 +/- 1.6%	89.0 +/- 1.2%	87.7 +/- 1.3%
Train: 20 (per class) Test: 131 (per class)				
f.40 -	78.9 +/- 1.6%	74.2 +/- 1.7%	89.1 +/- 1.2%	88.7 +/- 1.2%
f.30 -	77.8 +/- 1.6%	75.3 +/- 1.7%	90.0 +/- 1.2%	86.4 +/- 1.3%
f.15 -	77.6 +/- 1.6%	75.6 +/- 1.6%	90.4 +/- 1.2%	87.9 +/- 1.2%
Train: 10 per class Test: 141 per class				
f.40 -	75.4 +/- 1.6%	70.8 +/- 1.7%	79.1 +/- 1.5%	83.0 +/- 1.4%
f.30 -	74.1 +/- 1.6%	71.1 +/- 1.7%	84.1 +/- 1.3%	80.9 +/- 1.5%
f.15 -	73.8 +/- 1.6%	70.3 +/- 1.7%	81.1 +/- 1.4%	79.4 +/- 1.5%
Train: 5 per class Test: 131 per class				
f.40 -	66.8 +/- 1.7%	66.9 +/- 1.7%	72.9 +/- 1.6%	69.3 +/- 1.7%
f.30 -	72.0 +/- 1.6%	65.0 +/- 1.7%	75.2 +/- 1.6%	74.9 +/- 1.6%
f.15 -	64.2 +/- 1.7%	68.3 +/- 1.7%	75.5 +/- 1.6%	68.7 +/- 1.7%

Tiling: Original
Resolution: 640x480
Selection: Best Quality

Classes:
Nkandla
Msinga

Table A.8: Experiments: Region (R001b)

Train: 100 (per class) Test: 51 (per class)	
	default
f.40 -	83.6 +/- 2.3%
f.30 -	84.2 +/- 2.2%
f.15 -	83.6 +/- 2.3%
Train: 130 (per class) Test 21: (per class)	
f.30 -	85.2 +/- 3.4%
Train: 90 (per class) Test: 61 (per class)	
f.30 -	83.5 +/- 2.1%
Train: 80 (per class) Test 71 (per class)	
f.30 -	84.8 +/- 1.9%
Train: 50 (per class) Test: 101 (per class)	
f.30 -	82.4 +/- 1.7%
Train: 30 (per class) Test: 121 (per class)	
f.30 -	80.3 +/- 1.6%
Train: 20 (per class) Test: 131 (per class)	
f.30 -	78.2 +/- 1.6%
Train: 10 (per class) Test: 141 (per class)	
f.30 -	71.2 +/- 1.7%
Train: 5 (per class) Test: 146 (per class)	
f.30 -	68.1 +/- 1.7%

Tiling: Original
Resolution: 320x240
Selection: Best Quality

Classes:
Nkandla
Msinga

Table A.9: Experiments: Region (R003b)

Train: 100 (per class) Test: 90 (per class)	
	-L -C
f.40 -	91.6 +/- 1.3%
f.30-	89.8 +/- 1.4%
f.15 -	90.9 +/- 1.3%
Train: 130 (per class) Test 60: (per class)	
f.30 -	90.3 +/- 1.7%
Train: 90 (per class) Test: 100 (per class)	
f.30 -	90.1 +/- 1.3%
Train: 80 (per class) Test 110 (per class)	
f.30 -	90.2 +/- 1.2%
Train: 50 (per class) Test: 140 (per class)	
f.30 -	88.4 +/- 1.2%
Train: 30 (per class) Test: 160 (per class)	
f.30 -	87.9 +/- 1.1%
Train: 20 (per class) Test: 170 (per class)	
f.30 -	85.1 +/- 1.2%
Train: 10 (per class) Test: 180 (per class)	
f.30 -	80.7 +/- 1.3%
Train: 5 (per class) Test: 185 (per class)	
f.30 -	69.8 +/- 1.5%

Tiling: Original
Resolution: 320x240
Selection: Dark Edges Included

Classes:
Nkandla
Msinga

Table A.10: Experiments: Region (R004b)

Train: 100 (per class) Test: 69 (per class)	
	-L -C
f.40 -	82.0 +/- 2.0%
f.30-	92.0 +/- 1.4%
f.15 -	92.7 +/- 1.4%
Train: 120 (per class) Test 49: (per class)	
f.30 -	92.8 +/- 1.6%
Train: 90 (per class) Test: 79 (per class)	
f.30 -	90.8 +/- 1.4%
Train: 50 (per class) Test 129 (per class)	
f.30 -	90.4 +/- 1.1%
Train: 30 (per class) Test: 139 (per class)	
f.30 -	88.5 +/- 1.2%
Train: 20 (per class) Test: 149 (per class)	
f.30 -	86.5 +/- 1.2%
Train: 10 (per class) Test: 159 (per class)	
f.30 -	80.5 +/- 1.4%
Train: 5 (per class) Test: 164 (per class)	
f.30 -	71.6 +/- 1.5%

Tiling: Original
Resolution: 320x240
Selection: Blurry Included

Classes:
Nkandla
Msinga

Table A.11: Experiments: Region (R001e)

Train: 100 (per class) Test: 504 (per class)	
	- L - C
f.40 -	87.2 +/- 0.7%
f.30 -	86.6 +/- 0.7%
f.15 -	86.3 +/- 0.7%
Train: 500 (per class) Test: 104 (per class)	
f.30 -	90.0 +/- 1.1%
Train: 400 (per class) Test: 204 (per class)	
f.30 -	90.3 +/- 0.9%
Train: 300 (per class) Test: 304 (per class)	
f.30 -	88.8 +/- 0.8%
Train: 200 (per class) Test: 404 (per class)	
f.30 -	88.3 +/- 0.7%
Train: 130 per class Test 474 per class	
f.30 -	87.9 +/- 0.7%
Train: 90 p(er class) Test: 514 (per class)	
f.30 -	
Train: 80 (per class) Test: 524 (per class)	
f.30 -	87.0 +/- 0.6%
Train: 50 (per class) Test: 554 (per class)	
f.30 -	85.3 +/- 0.7%
Train: 30 (per class) Test: 574 (per class)	
f.30 -	83.3 +/- 0.7%
Train: 20 (per class) Test: 584 (per class)	
f.30 -	86.5 +/- 1.2%
Train: 10 (per class) Test: 594 (per class)	
f.30 -	76.7 +/- 0.8%
Train: 5 (per class) Test: 599 (per class)	
f.30 -	64.3 +/- 0.9%

Tiling: / 4
Resolutions: 320x240
Selection: Best Quality

Classes:
Nkandla
Msinga

Table A.12: Experiments: Region (R001c)

Train: 1500 (per class) Test: 916 (per class)	
	- L
f.30 -	77.3 +/- 0.6%
Train: 1000 (per class) Test: 1416 (per class)	
f.30 -	77.4 +/- 0.5%
Train: 500 (per class) Test: 1916 (per class)	
f.30 -	75.7 +/- 0.4%
Train: 100 (per class) Test: 2316 (per class)	
f.30 -	72.0 +/- 0.2%
Train: 50 per class Test 2366 per class	
f.30 -	69.6 +/- 0.4%
Train: 20 p(er class) Test: 2396 (per class)	
f.30 -	66.3 +/- 0.4%
Train: 10 (per class) Test: 2406 (per class)	
f.30 -	63.2 +/- 0.4%
Train: 5 (per class) Test: 2411 (per class)	
f.30 -	59.6 +/- 0.4%

Resolutions: 320x240
Tiling: / 15
Selection: Best Quality

Classes:
Nkandla
Msinga

Table A.13: Experiments: Community (C001a 2)

Train: 90 per class,Test 33 per class			
default	**-L -C**	**-W -L -C**	
f.40 -	84.3 +/- 3.3%	94.6 +/- 2.1%	95.7 +/- 1.9%
f.35 -	89.8 +/- 2.8%	96.1 +/- 1.8%	96.1 +/- 1.8%
f.30 -	88.3 +/- 2.9%	94.6 +/- 2.1%	95.9 +/- 1.8%
f.25 -	85.2 +/- 3.2%	94.8 +/- 2.0%	95.9 +/- 1.8%
f.20 -	86.3 +/- 3.1%	96.1 +/- 1.8%	96.1 +/- 1.8%
f.15 -	86.7 +/- 3.1%	96.7 +/- 1.6%	94.8 +/- 2.0%
f.10 -	84.3 +/- 3.3%	96.1 +/- 1.8%	96.1 +/- 1.8%
f.05-	83.3 +/- 3.4%	97.0 +/- 1.6%	95.2 +/- 2.0%
f.02-	73.7 +/- 4.0%	94.1 +/- 2.1%	97.0 +/- 1.6%
Train: 90 per class Test 33 per class			
f.40 -	86.1 +/- 2.6%	97.0 +/- 1.3%	97.0 +/- 1.3%
f.30 -	87.4 +/- 2.5%	95.6 +/- 1.6%	95.2 +/- 1.6%
f.15 -	85.3 +/- 2.7%	97.9 +/- 1.1%	95.9 +/- 1.5%
Train: 80 per class Test 43 per class			
f.40 -	84.8 +/- 2.4%	95.8 +/- 1.3%	94.9 +/- 1.5%
f.30 -	85.3 +/- 2.4%	96.3 +/- 1.3%	95.8 +/- 1.3%
f.15 -	**86.0 +/- 2.3%**	**96.6 +/- 1.2%**	**95.0 +/- 1.5%**
Train: 50 per class Test 73 per class			
f.40 -	83.2 +/- 1.9%	95.8 +/- 1.0%	95.9 +/- 1.0%
f.30 -	84.7 +/- 1.8%	94.2 +/- 1.2%	95.6 +/- 1.1%
f.15 -	82.7 +/- 1.9%	94.3 +/- 1.2%	95.1 +/- 1.1%
Train: 30 per class Test 93 per class			
f.40 -	81.9 +/- 1.7%	93.1 +/- 1.2%	94.4 +/- 1.0%
f.30 -	81.9 +/- 1.8%	94.6 +/- 1.0%	94.1 +/- 1.1%
f.15 -	81.7 +/- 1.8%	94.5 +/- 1.0%	94.2 +/- 1.1%
Train: 20 per class Test 103 per class			
f.40 -	81.2 +/- 1.7%	93.4 +/- 1.1%	92.9 +/- 1.1%
f.30 -	80.8 +/- 1.7%	92.4 +/- 1.1%	91.7 +/- 1.2%
f.15 -	80.4 +/- 1.7%	92.4 +/- 1.1%	93.0 +/- 1.1%
Train: 10 per class Test 113 per class			
f.40 -	74.6 +/- 1.8%	87.7 +/- 1.4%	88.7 +/- 1.3%
f.30 -	76.5 +/- 1.7%	89.2 +/- 1.3%	90.1 +/- 1.2%
f.15 -	70.0 +/- 1.9%	89.2 +/- 1.3%	89.7 +/- 1.3%
Train: 5 per class Test 118 per class			
f.40 -	70.3 +/- 1.8%	63.5 +/- 1.9%	81.7 +/- 1.6%
f.30 -	72.3 +/- 1.8%	73.9 +/- 1.8%	73.1 +/- 1.8%
f.15 -	65.3 +/- 1.9%	66.9 +/- 1.9%	80.8 +/- 1.6%

Resolutions: 640x480
Tiling: Original
Selection: Best Quality

Classes:
Mabaso
Mchuno

Table A.14: Experiments: Community (C001d 3)

Train: 100 per class Test 392 per class	
	- L - C
f.40 -	86.0 +/- 0.6%
f.35 -	85.1 +/- 0.6%
f.30 -	85.2 +/- 0.6%
f.25 -	85.5 +/- 0.6%
f.20 -	85.3 +/- 0.6%
f.15 -	85.7 +/- 0.6%
f.10 -	84.9 +/- 0.6%
f.05-	83.6 +/- 0.7%
f.02-	82.7 +/- 0.7%
Train: 90 per class Test 402 per class	
f.40 -	84.8 +/- 0.6%
f.30 -	85.2 +/- 0.6%
f.15 -	85.1 +/- 0.6%
Train: 80 per class Test 71 per class	
f.40 -	84.9 +/- 0.6%
f.30 -	84.3 +/- 0.6%
f.15 -	84.8 +/- 0.6%
Train: 50 per class Test 442 per class	
f.40 -	82.7 +/- 0.6%
f.30 -	83.3 +/- 0.6%
f.15 -	82.6 +/- 0.6%
Train: 30 per class Test 462 per class	
f.40 -	81.8 +/- 0.6%
f.30 -	79.1 +/- 0.7%
f.15 -	80.6 +/- 0.7%
Train: 20 per class Test 472 per class	
f.40 -	77.7 +/- 0.7%
f.30 -	76.7 +/- 0.7%
f.15 -	78.5 +/- 0.7%
Train: 10 per class Test 141 per class	
f.40 -	73.3 +/- 0.7%
f.30 -	72.0 +/- 0.7%
f.15 -	73.5 +/- 0.7%
Train: 5 per class Test 478 per class	
f.40 -	60.0 +/- 0.8%
f.30 -	65.8 +/- 0.8%
f.15 -	62.0 +/- 0.8%
Train: 400 per class Test 92 per class	
f.40 -	88.8 +/- 1.2%
f.30 -	89.2 +/- 1.2%
f.15 -	88.6 +/- 1.2%

Resolutions: 640x480
Tiling: /4
Selection: Best Quality

Classes:
Magwaza
Mabaso
Mchuno

Table A.15: Experiments: Community (C001a 3)

Train: 100 per class Test 23 per class			
default	**-L -C**	**-W -L -C**	
f.40 -	80.0 +/- 3.0%	93.9 +/- 1.8%	92.2 +/- 2.0%
f.35 -	79.1 +/- 3.0%	94.9 +/- 1.6%	92.0 +/- 2.0%
f.30 -	79.0 +/- 3.0%	93.3 +/- 1.9%	92.5 +/- 2.0%
f.25 -	76.4 +/- 3.2%	94.8 +/- 1.7%	92.8 +/- 1.9%
f.20 -	77.1 +/- 3.1%	94.2 +/- 1.7%	94.9 +/- 1.6%
f.15 -	73.9 +/- 3.3%	94.6 +/- 1.7%	92.8 +/- 1.9%
f.10 -	74.2 +/- 3.3%	94.3 +/- 1.7%	93.3 +/- 1.9%
f.05-	74.6 +/- 3.2%	93.5 +/- 1.8%	91.6 +/- 2.1%
f.02-	71.2 +/- 3.4%	92.0 +/- 2.0%	91.3 +/- 2.1%
Train: 90 per class Test 33 per class			
f.40 -	78.5 +/- 2.6%	92.3 +/- 1.7%	91.7 +/- 1.7%
f.30 -	75.5 +/- 2.7%	95.3 +/- 1.3%	93.0 +/- 1.6%
f.15 -	74.0 +/- 2.7%	94.2 +/- 1.5%	91.7 +/- 1.7%
Train: 80 per class Test 43 per class			
f.40 -	76.0 +/- 2.3%	92.7 +/- 1.4%	91.9 +/- 1.5%
f.30 -	77.4 +/- 2.3%	94.0 +/- 1.3%	91.2 +/- 1.5%
f.15 -	75.3 +/- 2.4%	94.3 +/- 1.3%	91.9 +/- 1.5%
Train: 50 per class Test 73 per class			
f.40 -	74.7 +/- 1.8%	91.0 +/- 1.2%	89.7 +/- 1.3%
f.30 -	75.9 +/- 1.8%	91.6 +/- 1.2%	89.9 +/- 1.3%
f.15 -	75.4 +/- 1.8%	91.5 +/- 1.2%	90.5 +/- 1.2%
Train: 30 per class Test 93 per class			
f.40 -	70.1 +/- 1.7%	88.6 +/- 1.2%	87.5 +/- 1.2%
f.30 -	70.5 +/- 1.7%	88.9 +/- 1.2%	87.4 +/- 1.2%
f.15 -	71.8 +/- 1.7%	88.1 +/- 1.2%	87.0 +/- 1.2%
Train: 20 per class Test 103 per class			
f.40 -	70.7 +/- 1.6%	87.3 +/- 1.2%	86.4 +/- 1.2%
f.30 -	69.9 +/- 1.6%	86.3 +/- 1.2%	85.3 +/- 1.2%
f.15 -	70.6 +/- 1.6%	85.1 +/- 1.3%	85.0 +/- 1.3%
Train: 10 per class Test 113 per class			
f.40 -	65.2 +/- 1.6%	80.0 +/- 1.3%	78.9 +/- 1.4%
f.30 -	66.7 +/- 1.6%	78.3 +/- 1.4%	79.9 +/- 1.3%
f.15 -	63.1 +/- 1.6%	80.0 +/- 1.3%	81.5 +/- 1.3%
Train: 5 per class Test 118 per class			
f.40 -	56.4 +/- 1.6%	68.2 +/- 1.5%	70.2 +/- 1.5%
f.30 -	54.4 +/- 1.6%	69.7 +/- 1.5%	71.2 +/- 1.5%
f.15 -	52.4 +/- 1.6%	69.0 +/- 1.5%	70.3 +/- 1.5%

Resolutions: 640x480
Tiling: Original
Selection: Best Quality

Classes:
Magwaza
Mabaso
Mchuno

Table A.16: Experiments: Vessel (V001a 5)

Train: 30 per class Test 11 per class	
	-L -C
f.40 -	74.1 +/- 2.6%
f.35 -	74.4 +/- 3.6%
f.30 -	74.6 +/- 2.6%
f.25 -	78.7 +/- 3.4%
f.20 -	77.1 +/- 3.5%
f.15 -	80.2 +/- 3.3%
f.10 -	77.1 +/- 3.5%
f.05-	74.7 +/- 3.6%
f.02-	74.7 +/- 3.6%
Train: 35 per class Test 6 per class	
f.40 -	80.7 +/- 4.5%
f.30 -	80.0 +/- 4.5%
f.15 -	80.2 +/- 3.2%
Train: 20 per class Test 21 per class	
f.40 -	76.6 +/- 2.6%
f.30 -	75.6 +/- 2.6%
f.15 -	75.5 +/- 2.6%
Train: 10 per class Test 31 per class	
f.40 -	73.3 +/- 2.2%
f.30 -	72.2 +/- 2.2%
f.15 -	73.2 +/- 2.2%
Train: 5 per class Test 36 per class	
f.40 -	70.1 +/- 2.1%
f.30 -	67.2 +/- 2.2%
f.15 -	68.6 +/- 2.1%

Resolutions: 640x480
Tiling: Original
Selection: Best Quality

Classes:
040
044
171
172
177

Table A.17: Experiments: Vessel (V001b5)

Train: 40 per class Test 18 per class	
	-L -C
f.40 -	90.2 +/- 2.9%
f.35 -	88.2 +/- 3.2%
f.30 -	89.0 +/- 3.1%
f.25 -	90.0 +/- 2.9%
f.20 -	88.0 +/- 3.2%
f.15 -	92.0 +/- 2.7%
f.10 -	86.8 +/- 3.3%
f.05-	87.8 +/- 3.2%
f.02-	80.0 +/- 3.9%
Train: 30 per class Test 18 per class	
f.40 -	88.3 +/- 2.1%
f.30 -	87.8 +/- 2.1%
f.15 -	87.7 +/- 2.1%
Train: 20 per class Test 28 per class	
f.40 -	86.1 +/- 1.8%
f.30 -	87.9 +/- 1.7%
f.15 -	85.6 +/- 1.8%
Train: 10 per class Test 38 per class	
f.40 -	80.6 +/- 1.8%
f.30 -	84.3 +/- 1.6%
f.15 -	82.8 +/- 1.7%
Train: 5 per class Test 43 per class	
f.40 -	74.7 +/- 1.8%
f.30 -	82.4 +/- 1.6%
f.15 -	77.6 +/- 1.8%

Resolutions: 320x240
Tiling: Original
Selection: Best Quality

Classes:
171
172
183
180
177

Table A.18: Experiments: Vessel (V001b 4)

Train: 40 per class Test 8 per class	
	-L -C
f.40 -	89.4 +/- 3.4%
f.35 -	88.1 +/- 3.5%
f.30 -	90.3 +/- 3.2%
f.25 -	90.0 +/- 3.3%
f.20 -	88.4 +/- 3.5%
f.15 -	86.6 +/- 3.7%
f.10 -	87.8 +/- 3.6%
f.05-	87.5 +/- 3.6%
f.02-	81.9 +/- 4.2%
Train: 30 per class Test 18 per class	
f.40 -	89.6 +/- 2.2%
f.30 -	87.1 +/- 2.4%
f.15 -	87.9 +/- 2.4%
Train: 20 per class Test 28 per class	
f.40 -	84.5 +/- 2.1%
f.30 -	84.7 +/- 2.1%
f.15 -	85.6 +/- 2.1%
Train: 10 per class Test 38 per class	
f.40 -	82.2 +/- 1.9%
f.30 -	82.1 +/- 1.9%
f.15 -	81.4 +/- 2.0%
Train: 5 per class Test 43 per class	
f.40 -	79.9 +/- 1.9%
f.30 -	70.1 +/- 2.2%
f.15 -	78.5 +/- 1.9%

Resolutions: 320x240
Tiling: Original
Selection: Best Quality

Classes:
171
172
183
180

Table A.19: Experiments: Vessel (V001b 3)

Train: 35 per class Test 6 per class	
	-L -C
f.30 -	91.1 +/- 4.2%
Train: 30 per class Test 11 per class	
f.30 -	85.5 +/- 3.8%
Train: 20 per class Test 21 per class	
f.30 -	80.6 +/- 3.1%
Train: 10 per class Test 31 per class	
f.30 -	79.2 +/- 2.6%
Train: 5 per class Test 36 per class	
f.30 -	70.3 +/- 2.7%

Resolution: 320x240
Tiling: Original
Selection: Best Quality

Classes:
044
040
177

Table A.20: Experiments: Vessel (V001b 8)

Train: 30 per class Test 5 per class	
	- L - C
f.30 -	77.0 +/- 4.1%
Train: 20 per class Test 15 per class	
f.30 -	76.3 +/- 2.4%
Train: 10 per class Test 25 per class	
f.30 -	70.3 +/- 2.0%
Train: 5 per class Test 30 per class	
f.30 -	67.8 +/- 1.9%

Resolution: 320x240
Tiling: Original
Selection: Best Quality

Classes:
040 172
044 177
045 180
171 183

Table A.21: Experiments: Shaping Technique (S001a)

Train: 150 per class Test 59 per class			
	default	**-L -C**	**-W**
f.40 -	66.8 +/- 2.0%	70.4 +/- 1.9%	61.4 +/- 2.0%
f.35 -	67.4 +/- 2.0%	71.9 +/- 1.9%	62.0 +/- 2.0%
f.30 -	67.6 +/- 2.0%	71.1 +/- 1.9%	60.1 +/- 2.1%
f.25 -	66.9 +/- 2.0%	71.8 +/- 1.9%	61.6 +/- 2.0%
f.20 -	66.9 +/- 2.0%	69.9 +/- 1.9%	61.6 +/- 2.0%
f.15 -	65.2 +/- 2.0%	71.1 +/- 1.9%	60.6 +/- 2.1%
f.10 -	67.4 +/- 2.0%	71.0 +/- 1.9%	61.9 +/- 2.0%
f.05 -	64.3 +/- 2.0%	68.7 +/- 1.9%	61.4 +/- 2.0%
f.02 -	62.1 +/- 2.0%	67.8 +/- 2.0%	59.8 +/- 2.1%
Train: 150 per class Test 59 per class			
f.40 -	66.3 +/- 2.7%	73.1 +/- 2.5%	68.3 +/- 2.7%
f.30 -	67.9 +/- 2.7%	71.4 +/- 2.6%	67.1 +/- 2.7%
f.15 -	67.5 +/- 2.7%	71.1 +/- 2.6%	65.3 +/- 2.7%
Train: 130 per class Test 79 per class			
f.40 -	66.3 +/- 2.3%	71.3 +/- 2.2%	67.8 +/- 2.3%
f.30 -	66.1 +/- 2.3%	70.8 +/- 2.2%	66.7 +/- 2.3%
f.15 -	67.3 +/- 2.3%	68.9 +/- 2.3%	64.4 +/- 2.4%
Train: 90 per class Test 119 per class			
f.40 -	67.1 +/- 1.9%	70.4 +/- 1.8%	66.8 +/- 1.9%
f.30 -	65.3 +/- 1.9%	71.4 +/- 1.8%	67.4 +/- 1.9%
f.15 -	66.8 +/- 1.9%	69.7 +/- 1.8%	65.3 +/- 1.9%
Train: 80 per class Test 129 per class			
f.40 -	64.1 +/- 1.9%	70.4 +/- 1.8%	65.9 +/- 1.8%
f.30 -	66.2 +/- 1.8%	70.6 +/- 1.8%	65.3 +/- 1.8%
f.15 -	67.0 +/- 1.8%	68.3 +/- 1.8%	
Train: 50 per class Test 159 per class			
f.40 -	65.2 +/- 1.7%	68.6 +/- 1.6%	65.4 +/- 1.7%
f.30 -	66.7 +/- 1.6%	68.8 +/- 1.6%	63.2 +/- 1.7%
f.15 -	65.2 +/- 1.7%	68.9 +/- 1.6%	63.8 +/- 1.7%
Train: 30 per class Test 179 per class			
f.40 -	63.2 +/- 1.6%	67.1 +/- 1.5%	64.6 +/- 1.6%
f.30 -	57.7 +/- 1.5%	69.3 +/- 1.5%	66.2 +/- 1.5%
f.15 -	64.5 +/- 1.6%	66.6 +/- 1.5%	62.8 +/- 1.6%
Train: 20 per class Test 189 per class			
f.40 -	66.1 +/- 1.5%	66.2 +/- 1.5%	63.8 +/- 1.5%
f.30 -	60.8 +/- 1.6%	65.3 +/- 1.5%	64.6 +/- 1.5%
f.15 -	64.7 +/- 1.5%	64.7 +/- 1.5%	64.1 +/- 1.5%
Train: 10 per class Test 199 per class			
f.40 -	62.1 +/- 1.5%	64.7 +/- 1.5%	58.3 +/- 1.5%
f.30 -	60.7 +/- 1.5%	61.0 +/- 1.5%	60.3 +/- 1.5%
f.15 -	58.6 +/- 1.5%	60.7 +/- 1.5%	62.6 +/- 1.5%
Train: 5 per class Test 204 per class			
f.30 -	54.5 +/- 1.5%	54.8 +/- 1.5%	59.2 +/- 1.5%

Resolutions: 640x480
Tiling: Original
Selection: Best Quality

Classes:
Bases
Walls

Table A.22: Experiments: Shaping Technique (S001d)

Train: 100 per class Test 736 per class	
	-L -C
f.40 -	60.9 +/- 0.8%
f.30 -	62.9 +/- 0.8%
f.15 -	62.2 +/- 0.8%
Train: 500 per class Test 336 per class	
f.30 -	64.3 +/- 1.1%
Train: 400 per class Test436 per class	
f.30 -	64.6 +/- 1.0%
Train: 300 per class Test 536 per class	
f.30 -	64.2 +/- 0.9%
Train: 200 per class Test 636 per class	
f.30 -	64.0 +/- 0.8%
Train: 130 per class Test 706 per class	
f.30 -	63.2 +/- 0.8%
Train: 90 per class Test 746 per class	
f.30 -	61.4 +/- 0.8%
Train: 80 per class Test 756 per class	
f.30 -	62.8 +/- 0.8%
Train: 50 per class Test 786 per class	
f.30 -	61.3 +/- 0.8%
Train: 30 per class Test 806 per class	
f.30 -	57.6 +/- 0.8%
Train: 20 per class Test 816 per class	
f.30 -	59.9 +/- 0.8%
Train: 10 per class Test 826 per class	
f.30 -	56.2 +/- 0.8%
Train: 5 per class Test 263 per class	
f.30 -	52.3 +/- 0.8%

Resolutions: 640x480
Tiling: /4
Selection: Best Quality

Classes:
Bases
Walls

Table A.23: Experiments: Potter (IC001a)

Train: 50 per class Test 17 per class					
	default	**-L**	**-C**	**-L -C**	**-W -L -C**
f.40 -	90.6 +/- 3.1%	94.7 +/- 2.4%	92.9 +/- 2.7%	95.0 +/- 2.3%	94.2 +/- 1.7%
f.35 -	92.4 +/- 2.8%	94.1 +/- 2.5%	94.1 +/- 2.5%	93.2 +/- 2.7%	94.6 +/- 1.6%
f.30 -	90.9 +/- 3.1%	95.0 +/- 2.3%	94.7 +/- 2.4%	95.6 +/- 2.2%	95.5 +/- 1.5%
f.25 -	92.1 +/- 2.9%	97.1 +/- 1.8%	96.2 +/- 2.0%	94.1 +/- 2.5%	95.0 +/- 1.6%
f.20 -	87.9 +/- 3.5%	95.9 +/- 2.1%	95.3 +/- 2.3%	93.8 +/- 2.6%	94.2 +/- 1.7%
f.15 -	90.0 +/- 3.2%	96.8 +/- 1.9%	92.4 +/- 2.8%	95.3 +/- 2.3%	95.7 +/- 1.5%
f.10 -	91.2 +/- 3.0%	93.8 +/- 2.6%	93.2 +/- 2.7%	93.8 +/- 2.6%	93.6 +/- 1.8%
f.05-	91.8 +/- 2.9%	96.2 +/- 2.0%	96.2 +/- 2.0%	93.2 +/- 2.7%	95.8 +/- 1.4%
f.02-	91.2 +/- 3.0%	94.1 +/- 2.5%	94.1 +/- 2.5%	93.2 +/- 2.7%	92.4 +/- 1.9%
Train: 30 per class Test 37 per class					
f.40 -	91.2 +/- 2.0%	93.4 +/- 1.8%	93.8 +/- 1.7%	95.3 +/- 1.5%	93.4 +/- 1.8%
f.30 -	90.0 +/- 2.2%	94.9 +/- 1.6%	94.6 +/- 1.6%	92.0 +/- 2.0%	93.9 +/- 1.7%
f.15 -	90.1 +/- 2.1%	93.9 +/- 1.7%	93.9 +/- 1.7%	92.2 +/- 1.9%	93.9 +/- 1.7%
Train: 20 per class Test 47 per class					
f.40 -	90.1 +/- 2.1%	93.1 +/- 1.6%	92.2 +/- 1.7%	93.8 +/- 1.5%	94.6 +/- 1.4%
f.30 -	91.7 +/- 1.8%	93.6 +/- 1.6%	95.4 +/- 1.3%	93.1 +/- 1.6%	94.0 +/- 1.5%
f.15 -	90.3 +/- 1.9%	91.7 +/- 1.8%	93.5 +/- 1.6%	93.6 +/- 1.6%	94.3 +/- 1.5%
Train: 10 per class Test 57 per class					
f.40 -	86.9 +/- 2.0%	91.4 +/- 1.6%	91.2 +/- 1.6%	91.3 +/- 1.6%	91.8 +/- 1.6%
f.30 -	87.3 +/- 1.9%	92.3 +/- 1.5%	92.5 +/- 1.5%	93.4 +/- 1.4%	93.5 +/- 1.4%
f.15 -	88.6 +/- 1.8%	92.1 +/- 1.6%	89.2 +/- 1.8%	92.2 +/- 1.6%	91.7 +/- 1.6%
Train: 5 per class Test 62 per class					
f.40 -	81.9 +/- 2.1%	76.1 +/- 2.4%	84.6 +/- 2.0%	77.7 +/- 2.3%	83.9 +/- 2.0%
f.30 -	77.7 +/- 2.3%	74.0 +/- 2.4%	81.3 +/- 2.2%	84.7 +/- 2.0%	87.9 +/- 1.8%
f.15 -	76.4 +/- 2.4%	79.4 +/- 2.2%	85.6 +/- 2.0%	82.7 +/- 2.1%	86.0 +/- 1.9%

Resolutions: 640x480
Tiling: Original
Selection: Best Quality

Classes: (Walls Only)
Potter A
Potter B

Table A.24: Experiments: Potter (IC001d)

Train: 200 per class Test 68 per class	
	- L -C
f.40 -	94.9 +/- 1.2%
f.30 -	94.5 +/- 1.2%
f.15 -	93.4 +/- 1.3%
Train: 130 per class Test 138 per class	
f.30 -	93.8 +/- 0.9%
Train: 100 per class Test 168 per class	
f.30 -	94.3 +/- 0.8%
Train: 50 per class Test 218 per class	
f.30 -	92.5 +/- 0.8%
Train: 30 per class Test 238 per class	
f.30 -	90.9 +/- 0.8%
Train: 20 per class Test 248 per class	
f.30 -	90.4 +/- 0.8%
Train: 10 per class Test 258 per class	
f.30 -	86.9 +/- 0.9%
Train: 5 per class Test 263 per class	
f.30 -	80.3 +/- 1.1%

Resolutions: 640x480
Tiling: /4
Selection: Best Quality

Classes: (Walls Only)
Potter A
Potter B

Table A.25: Experiments: Potter (IC001b)

Train: 50 per class Test 17 per class	
	- L - C
f.40 -	95.6 +/- 2.2%
f.30 -	94.7 +/- 2.4%
f.15 -	96.8 +/- 1.9%
Train: 30 per class Test 37 per class	
f.30 -	93.6 +/- 1.8%
Train: 20 per class Test 47 per class	
f.30 -	93.8 +/- 1.5%
Train: 10 per class Test 57 per class	
f.30 -	91.1 +/- 1.6%
Train: 5 per class Test 62 per class	
f.30 -	79.2 +/- 2.3%

Resolutions: 320x240
Tiling: Original
Selection: Best Quality

Classes: (Walls only)
Potter A
Potter B

Table A.26: Experiments: Potter (IC003b)

Train: 50 per class Test 44 per class	
	-L -C
f.40 -	94.5 +/- 1.5%
f.30 -	94.1 +/- 1.6%
f.15 -	94.5 +/- 1.5%
Train: 30 per class Test 64 per class	
f.30 -	94.8 +/- 1.2%
Train: 20 per class Test 74 per class	
f.30 -	92.0 +/- 1.4%
Train: 10 per class Test 84 per class	
f.30 -	89.3 +/- 1.5%
Train: 5 per class Test 89 per class	
f.30 -	77.8 +/- 1.9%

Resolutions: 320x240
Tiling: Original
Selection: Dark Edges Incl.

Classes:(Walls only)
Potter A
Potter B

Table A.27: Experiments: Potter (IC001b 3)

Train: 40 per class Test 24 per class	
	- L - C
f.30 -	94.7 +/- 1.6%
Train: 35 per class Test 29 per class	
f.30 -	96.8 +/- 1.2%
Train: 30 per class Test 34 per class	
f.30 -	95.5 +/- 1.3%
Train: 25 per class Test 39 per class	
f.30 -	95.1 +/- 1.2%
Train: 20 per class Test 44 per class	
f.30 -	94.6 +/- 1.2%
Train: 10 per class Test 36 per class	
f.30 -	93.3 +/- 1.2%
Train: 5 per class Test 59 per class	
f.30 -	83.5 +/- 1.7%

Resolution: 320x240
Split: Original
Selection: Best Quality

Classes: (Walls only)
Potter A
Potter B
Potter C

Table A.28: Experiments: Potter (IC004b)

Train: 50 per class Test 20 per class	
	-L -C
f.40 -	94.5 +/- 2.2%
f.30 -	93.2 +/- 2.5%
f.15 -	95.8 +/- 2.0%
Train: 30 per class Test 40 per class	
f.30 -	93.8 +/- 1.7%
Train: 20 per class Test 50 per class	
f.30 -	92.9 +/- 1.6%
Train: 10 per class Test 60 per class	
f.30 -	91.6 +/- 1.6%
Train: 5 per class Test 65 per class	
f.30 -	77.5 +/- 2.3%

Resolutions: 320x240
Tiling: Original
Selection: Blurry Incl.

Classes:(Walls only)
Potter A
Potter B

Table A.29: Experiments: Potter (IS001a)

Train: 45 per class Test 45 per class					
	default	**-L**	**-C**	**-L -C**	**-W -L -C**
f.40 -	90.0 +/- 5.9%	91.0 +/- 5.6%	93.0 +/- 5.0%	95.0 +/- 4.3%	95.3 +/- 3.7%
f.35 -	95.0 +/- 4.3%	94.0 +/- 4.7%	88.0 +/- 6.4%	93.0 +/- 5.0%	94.0 +/- 4.7%
f.30 -	88.0 +/- 6.4%	94.0 +/- 4.7%	90.0 +/- 5.9%	95.0 +/- 4.3%	87.0 +/- 6.6%
f.25 -	89.0 +/- 6.1%	94.0 +/- 4.7%	92.0 +/- 5.3%	95.0 +/- 4.3%	95.0 +/- 4.3%
f.20 -	88.0 +/- 6.4%	91.0 +/- 5.6%	94.0 +/- 4.7%	94.3 +/- 4.1%	92.0 +/- 5.3%
f.15 -	92.0 +/- 5.3%	94.0 +/- 4.7%	87.0 +/- 6.6%	93.0 +/- 5.0%	94.0 +/- 4.7%
f.10 -	90.0 +/- 5.9%	92.0 +/- 5.3%	92.0 +/- 5.3%	94.3 +/- 4.1%	93.0 +/- 5.0%
f.05-	86.0 +/- 6.8%	90.0 +/- 5.9%	85.0 +/- 7.0%	94.0 +/- 4.7%	88.0 +/- 6.4%
f.02-	90.0 +/- 5.9%	86.0 +/- 6.8%	91.0 +/- 5.6%	88.0 +/- 6.4%	87.0 +/- 6.6%
Train: 30 per class Test 20 per class					
f.40 -	86.5 +/- 3.3%	91.0 +/- 2.8%	88.8 +/- 3.1%	93.2 +/- 2.5%	94.8 +/- 2.2%
f.30 -	88.2 +/- 3.2%	92.5 +/- 2.6%	89.8 +/- 3.0%	93.2 +/- 2.5%	90.8 +/- 2.8%
f.15 -	90.5 +/- 2.9%	89.8 +/- 3.0%	91.8 +/- 2.7%	94.2 +/- 2.3%	92.0 +/- 2.7%
Train: 20 per class Test 30 per class					
f.40 -	87.0 +/- 2.7%	92.5 +/- 2.1%	89.2 +/- 2.5%	92.7 +/- 2.1%	91.7 +/- 2.2%
f.30 -	89.2 +/- 2.5%	90.3 +/- 2.4%	90.3 +/- 2.4%	93.3 +/- 2.0%	92.3 +/- 2.1%
f.15 -	86.2 +/- 2.8%	88.2 +/- 2.6%	90.0 +/- 2.4%	90.3 +/- 2.4%	92.2 +/- 2.1%
Train: 10 per class Test 40 per class					
f.40 -	87.1 +/- 2.3%	86.2 +/- 2.4%	89.6 +/- 2.1%	89.2 +/- 2.1%	90.4 +/- 2.0%
f.30 -	87.5 +/- 2.3%	89.8 +/- 2.1%	90.6 +/- 2.0%	89.1 +/- 2.2%	90.4 +/- 2.0%
f.15 -	86.1 +/- 2.4%	89.1 +/- 2.2%	89.4 +/- 2.1%	91.0 +/- 2.0%	91.5 +/- 1.9%
Train: 5 per class Test 45 per class					
f.40 -	76.6 +/- 2.8%	82.4 +/- 2.5%	84.7 +/- 2.4%	76.1 +/- 2.8%	82.0 +/- 2.5%
f.30 -	85.8 +/- 2.3%	84.4 +/- 2.4%	82.4 +/- 2.5%	77.4 +/- 2.7%	82.3 +/- 2.5%
f.15 -	74.8 +/- 2.8%	78.2 +/- 2.7%	79.3 +/- 2.6%	82.2 +/- 2.5%	85.9 +/- 2.3%

Resolution: 640x480
Tiling: /4
Selection: Best Quality

Classes: (Bases only)
Potter A
Potter B

Table A.30: Experiments: Potter (IS001e)

Train: 100 per class Test 100 per class	
	-L -C
f.40 -	91.1 +/- 1.2%
f.30 -	91.8 +/- 1.2%
f.15 -	91.1 +/- 1.2%
Train: 180 per class Test 20 per class	
f.30 -	93.2 +/- 2.5%
Train: 150 per class Test 50 per class	
f.30 -	90.2 +/- 1.8%
Train: 130 per class Test 218 per class	
f.30 -	91.0 +/- 1.5%
Train: 90 per class Test 110 per class	
f.30 -	91.0 +/- 1.2%
Train: 50 per class Test 150 per class	
f.30 -	90.6 +/- 1.0%
Train: 30 per class Test 170 per class	
f.30 -	89.6 +/- 1.0%
Train: 20 per class Test 180 per class	
f.30 -	87.6 +/- 1.1%
Train: 10 per class Test 190 per class	
f.30 -	85.3 +/- 1.1%
Train: 5 per class Test 195 per class	
f.30 -	78.2 +/- 1.3%

Resolution: 640x480
Tiling: /4
Selection: Best Quality

Classes: (Bases only)
Potter A
Potter B

Table A.31: Experiments: Potter (IS001b)

Train: 45 per class Test 5 per class	
	-L -C
f.40 -	95.0 +/- 4.3%
f.30 -	93.0 +/- 5.0%
f.15 -	94.0 +/- 4.7%
Train: 30 per class Test 20 per class	
f.30 -	93.0 +/- 2.5%
Train: 20 per class Test 30 per class	
f.30 -	92.7 +/- 2.1%
Train: 10 per class Test 40 per class	
f.30 -	89.9 +/- 2.1%
Train: 5 per class Test 45 per class	
f.30 -	74.4 +/- 2.8%

Resolution: 320x240
Tiling: Original
Selection: Best Quality

Classes: (Bases only)
Potter A
Potter B

Table A.32: Experiments: Potter (IBW001b 3)

Train: 50 per class Test 16 per class	
	-L -C
f.40 -	95.4 +/- 1.9%
f.35 -	95.0 +/- 1.9%
f.30 -	97.5 +/- 1.4%
f.25 -	95.0 +/- 1.9%
f.20 -	96.7 +/- 1.6%
f.15 -	96.2 +/- 1.7%
f.10 -	94.4 +/- 2.1%
f.05-	94.8 +/- 2.0%
f.02-	91.5 +/- 2.5%
Train: 60 per class Test 6 per class	
f.40 -	96.7 +/- 2.6%
f.30 -	96.1 +/- 2.8%
f.15 -	93.9 +/- 3.5%
Train: 30 per class Test 36 per class	
f.40 -	95.0 +/- 1.3%
f.30 -	94.4 +/- 1.4%
f.15 -	94.6 +/- 1.3%
Train: 20 per class Test 46 per class	
f.40 -	93.8 +/- 1.3%
f.30 -	93.8 +/- 1.3%
f.15 -	91.9 +/- 1.4%
Train: 10 per class Test 61 per class	
f.40 -	91.5 +/- 1.3%
f.30 -	90.1 +/- 1.4%
f.15 -	91.2 +/- 1.4%
Train: 5 per class Test 61 per class	
f.40 -	85.8 +/- 1.6%
f.30 -	87.2 +/- 1.5%
f.15 -	81.1 +/- 1.8%

Resolution: 320x240
Tiling: Original
Selection: Best Quality

Classes: (Walls and Bases)
Potter A
Potter B
Potter C

Table A.33: Experiments: Potter (IBW001b 2)

Train: 50 per class Test 73 per class	
	-L -C
f.40 -	94.7 +/- 1.2%
f.35 -	94.4 +/- 1.2%
f.30 -	93.7 +/- 1.2%
f.25 -	94.5 +/- 1.2%
f.20 -	94.6 +/- 1.2%
f.15 -	94.2 +/- 1.2%
f.10 -	94.4 +/- 1.2%
f.05-	94.2 +/- 1.2%
f.02-	92.9 +/- 1.3%
Train: 60 per class Test 63 per class	
f.40 -	94.1 +/- 1.3%
f.30 -	94.1 +/- 1.3%
f.15 -	94.4 +/- 1.3%
Train: 30 per class Test 93 per class	
f.40 -	93.9 +/- 1.1%
f.30 -	94.2 +/- 1.1%
f.15 -	93.9 +/- 1.1%
Train: 20 per class Test 103 per class	
f.40 -	93.4 +/- 1.1%
f.30 -	93.2 +/- 1.1%
f.15 -	91.7 +/- 1.2%
Train: 10 per class Test 113 per class	
f.40 -	89.2 +/- 1.3%
f.30 -	89.5 +/- 1.3%
f.15 -	91.7 +/- 1.2%
Train: 5 per class Test 118 per class	
f.40 -	81.1 +/- 1.6%
f.30 -	80.0 +/- 1.6%
f.15 -	86.4 +/- 1.4%

Resolution: 320x240
Tiling: Original
Selection: Best Qaulity

Classes: (Walls and Bases)
Potter A
Potter B

Table A.34: Experiments: Potter (IBW001b 4)

Train: 35 per class Test 6 per class	
	-L -C
f.30 -	91.7 +/- 3.5%
Train: 30 per class Test 11 per class	
f.30 -	92.5 +/- 2.5%
Train: 20 per class Test 21 per class	
f.30 -	89.5 +/- 2.1%
Train: 31 per class Test 10 per class	
f.30 -	87.6 +/- 1.8%
Train: 5 per class Test 36 per class	
f.30 -	84.1 +/- 1.9%

Resolution: 320x240
Tiling: Original
Selection: Best Quality

Classes:(Bases and Walls)
Potter A
Potter B
Potter C
Potter D

Appendix B

Highest Weighted Features by Classification Problem

Table B.1: Highest Weighted Features

Dataset	**Description**	**Highest Weighted Feature Categories**
Region (R001a)	Classes: Msinga, Nkandla	Gini Coefficient [Hue] Zernike Coefficient [Hue] Heralick Textures [Wavelet / Fourier] Comb Moments [Hue] Multiscale Histogram Tamura Textures [Fourier Hue] Pixel Intensity Statistics [Hue]
Community (C001a_2)	Classes: Mabaso, Mchuno	Heralick Textures [Hue]

Table B.1: Highest Weighted Features

Dataset	**Description**	**Highest Weighted Feature Categories**
Community (C001a_3)	Classes: Magwaza, Mabaso, Mchuno	Heralick Textures [Fourier / Wavelet] Heralick Textures [Hue] Gini Coefficient [Hue]
Vessels (V001a_5)	Classes: 040, 044, 171, 172, 177	Multiscale Histogram [Hue] Heralick Textures [Hue] Pixel Intensity [Fourier Hue] Pixel Intensiy [Hue] Radon Coeficient [Fouries / Hue]
Vessels (V001b_5)	Classes: 171, 172, 177, 183, 180	Heralick Textures [Hue] Pixel Intensity Statistics [Chebyshev / Hue] Pixel intensity Statistics [Hue] Fractal Features [Chebyshev Hue]
Vessels: (V001b_4)	Classes: 171, 172, 183, 180	Fractal Features [Colour Transform] Fractal Features Multiscale Histogram [Hue] Heralick textures [Hue]

Table B.1: Highest Weighted Features

Dataset	**Description**	**Highest Weighted Feature Categories**
Vessel (V001b_3)	Classes: 044, 040, 177	Colour Histogram Heralick Texture [Hue] Pixel Intensity Statistics [Hue] Fractal Features [Chebyshev / Hue] Tamura Textures [Hue]
Vessel (V001b_8)	Classes: 040, 044, 045, 171, 172, 177, 180, 183	Pixel Intensity Statistics [Fourier / Hue] Pixel Intensity Statistics [Chebyshev] Heralick Textures [Hue] Multiscale Histogram [Hue] Fractal Features [Chebyshebv \Hue]
Shaping (S001a)	Classes: Bases, Walls (all vessels included)	Heralick Textures [Colour Transform] Tamura Textures [Chebyshev] Multiscale Histograms [Chebyshev] Chebyshev Coefficient [Colour Transform]
Individual (IS001a)	Classes: Potters A, B (Bases only)	Heralick Textures [Fourier / Wavelet] Fractal Features Fractal Features [Colour Transfrom]

Table B.1: Highest Weighted Features

Dataset	**Description**	**Highest Weighted Feature Categories**
Individual (IC001a)	Classes: Potters A, B (Walls only)	Heralick Textures [Colour Transform] Heralick Textures Heralic Textures [Fourier / Wavelet] Tamura Textures
Individual (IC001a_3)	Classes: Potters A, B, C (Walls only)	Pixel Intensity Statistics [Hue] Radon Coefficient [Fourier / Hue] Fractal Features [Fourier / Hue] Fractal Features [Chebyshev / Hue]
Individual (IBW001b_2)	Classes: Potters A, B (Walls and Bases)	Heralick Textures [Colour Transform] Heralick Textures Heralick Textures [Fourier / Wavelet] Fractal Features Fractal Features [Colour Transform]
Individual (IBW001b_3)	Classes: Potters A, B, C (Walls and Bases)	Pixel Statistics [Hue] Pixel Intensity Statistics [Chebyshev / Hue] Fractal Features [Chebyshev / Hue] Fractal Features [Hue]

Table B.1: Highest Weighted Features

Dataset	**Description**	**Highest Weighted Feature Categories**
Individual (IBW001b_4)	Classes: Potters A, B, C, D (Walls and Bases)	Pixel Intensity Statistics [Fourier / Hue] Pixel Intensity Statistics [Chebyshev / Hue] Fractal Features [Chebyshev / Hue] Multiscale Histogram [Hue]

Bibliography

Adams, W. & Adams, E. (1991). *Archaeological typology and practical reality. a deialectical approach to artifact classification and sorting*. Cambridge: Cambridge University Press.

Arnold, D. (1974). Some principles of paste analysis and interpretation: A preliminary formulation. *Journal of the Steward Anthropological Society*, (6), 33–47.

Balfet, H. & Matson, F. (1965). Ethnographical observations in north africa and archaeological interpretation: The pottery of the maghreb. In H. Lechtman & R. Merrill (Eds.), *Material culture: Styles, organization, and dynamics of technology* (pp. 161–177). New York: Viking Fund Publications in Anthropology No. 41.

Banning, E. (2000). *The archaeologist's laboratory. the analysis of archaelogical data*. New York: Kluwer Academic.

Berg, I. (2008). Looking through pots: Recent advances in ceramic x-radiography. *Journal of Archaeological Science*, (35), 1177–1188.

Bickler, S. (2018). Machine learning identification and classification of historic ceramics. *Archaeology in New Zealand*, (61), 48–58.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Cambridge: Cambridge University Press.

Blackman, M. J., Stein, G. J., & Vandiver, P. B. (1993). The standardization hypothesis and ceramic mass production: Technological, compositional,and metric indexes of craft specialization at tell leilan, syria. *American Antiquity*, (58), 60–80.

Blandino, B. (2003). *Copiled pottery*. London: A&C Black.

Carmichael, P. (1986). Nasca pottery construction. *Nawpa Pacha: Journal of Andean Archaeology*, (24), 31–48.

Courty, M. (1994). Identification of wheel throwing on the basis of ceramic surface features and microfabrics. *Journal of Archaeological Science*, (22), 17–50.

Debroutelle, T., Treuillet, S., Chetouani, A., Exbrayat, M., & Jesset, S. (2017). Automatic classification of ceramic sherds with relief motifs. *Journal of Electronic Imaging*, (2).

Dunnell, R. (1971). *Systematics in prehistory*. New York: The Free Press.

Ford, J. & Steward, J. (1954). On the concept of types. *Americal Anthropologist*, (56), 42–57.

Fowler, K. [K.], Middleton, E., & Fayek, M. (2017). The human element: Discerning the effects of potter's behavior on the chemical composition of ceramics. *Archaeological and Anthropological Sciences*, 1–28.

Fowler, K. [K.D.]. (2006). Classification and collapse: The ethnohistory of zulu ceramic use. *Southern African Humanities*, (18), 93–117.

Fowler, K. [K.D.]. (2008). Zulu pottery production in the lower thukela basin, kwazulu-natal, south africa. *Southern African Humanities*, (20), 477–511.

Fowler, K. [K.D.]. (2011). The zulu ceramic tradition in msinga, south africa. *Southern African Humanities*, (23), 173–202.

Gabor, D. (1994). Theory of communication. *Journal of IEEE*, (93), 429–457.

Gradshtein, I. & Ryzhik, I. (1994). *Table of integrals, series and products 5th edition*. Academic Press.

Gray, S. (1971). Local properties of binary images in two dimensions. *IEEE Trans on Computers*, (20), 55–56.

Gregorescu, C., Petkov, N., & Kruizinga, P. (2002). Comparison of texture features based on gabor filters. *IEEE Trans on Image Processing*, (11), 160–167.

Hadjidementriou, E., Grossberg, M., & Nayar, C. (2001). Spacial information in multiresolution histograms. *IEEE Conference on Computer Vision and Pattern Recognition*, 1–702.

Hand, D. (1997). *Construction and assessement of classification rules*. New York: John Wiley and Sons.

Hangstrum, M. (1985). Measuring prehistoric ceramic craft specialization: A test case in the american southwest. *Journal of Field Archaeology*, (12), 65–75.

Haralick, R., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Trans on Image Processing*, (6), 269–285.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning. data mining, inference, and predition*. New York: Springer.

Hill, D. (1984). An analysis of voids in ceramics. *Lambda Alpha Journal*, (16), 1–28.

Hörr, C., Brunner, D., & Brunnett, G. (2007). Feature extraction on axially symmetric pottery for hierarchical classification. *Computer-Aided Design and Applications*, (4), 375–384.

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

Kahl, W. & Ramminger, B. (2012). Non-destructive fabric analysis of prehistoric pottery using high-resolution x-ray microtomography: A pilot study on the late mesolithic to neolithic site hamburg-boberg. *Journal of Archaeological Science*, (39), 2206–2219.

Kohavi, R. & Provost, F. (1998). Glossary of terms. *Machine Learning*, *30*(2-3), 271–274.

Kuminski, E., George, J., Wallin, J., & Shamir, L. (2014). Combining human and machine learning for morphological analysis of galaxy images. *Publications of the Astronomical Society of the Pacific*, (126), 959–967.

Lechtman, H. (1977). Style in technology - some early thoughts. In H. Lechtman & R. Merrill (Eds.), *Material culture: Styles, organization, and dynamics of technology* (pp. 3–20). New York: West Publishing.

Lim, J. (1990). *Two-dimensional signal and image processing*. New Haven: Prentice Hall.

Lindahl, A. & Pikirayi, I. (2010). Ceramics and change: An overview of pottery production techniques in northern south africa and eastern zimbabwe during the first and second millennium ad. *Archaeol. Anthropol. Sci.*

Mackay, J. (2003). *Information theory, inference, and learning algorithms*. Cambridge: Cambridge University Press.

Makridis, M. & Daras, P. (2012). Automatic classification of archaeological pottery sherds. *Journal on Computing and Cultural Heritage*, (5), 1–21.

McGovern, P. (1989). Ancient ceramic technology and stylistic change: Contrasting studies from southwest asia. In J. Henderson (Ed.), *Scientific analysis in archaeology and its interpretation* (pp. 63–81). Oxford: Oxford University Press.

Murphy, R., Velliste, M., Yao, J., & Porreca, G. (2001). Searching online journals for flurescence microscopy images depicting protein subcellular location patterns. *Proc 2nd IEEE International Symposium on Bioinformatics and Biomedical Engineering*, 19–128.

Nguifo, E., Lagrange, M., Renaud, M., & Sallantin, J. (1997). Plata: An application of legal, a machine learning based system, to a typology of archaeological ceramics. *Computers and the Humanities*, (31), 169–87.

Orlov, N., Johnston, J., Macura, T., Shamir, L., & Goldberg, I. (2007). Computer vision for microscopy applications. In G. Obinata & A. Dutta (Eds.), *Vision systems - segmentation and pattern recognition* (pp. 221–242). Vienna: ARS Press.

Orlov, N., Shamir, L., Macura, T., Johnston, J., Eckley, M., & Goldberg, I. (2008). Wnd-charm: Multi-purpose image classification using compound image transforms. *Pattern Recognition Letters*, (29), 1684–1693.

Orton, C. & Hughes, M. (2013). *Pottery in archaeology*. Cambridge: Cambridge University Press.

Otsu, N. (1979). A threshold selection method from gray level histograms. *IEEE Trans Systems, Man and Cybernetics*, (9), 62–66.

Prewitt, J. (1970). Object enhancemebt and extraction. In B. Lipkin & A. Rosenfeld (Eds.), *Picture processing psychopictoris* (pp. 75–149). New York: New York: Academic.

Rice, P. (1987). *Pottery analysis: A sourcebook*. Chicago: University of Chicago Press.

Ross, J., Fowler, K., Shai, I., Greenfield, H., & Maeir, A. (2018). A scanning method for the identification of pottery forming techniques at the mesoscopic scale: A pilot study in the manufacture of early bronze age iii holemouth jars and platters from tell es-safi/gath. *Journal of Archeological Science*, (18), 551–561.

Rye, O. (1977). Pottery manufacturing techniques: X-ray studies. *Archaeonometry*, (19), 205–211.

Rye, O. (1981). *Pottery technology. principles and reconstruction.* Washington: Taraxacum.

Samuel, A. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *3*(3).

Shai, S. & Shai, B. (2014). *Understanding machine learning. from theory to algorithms*. Cambridge: Cambridge University Press.

Shamir, L. (2011). Pose and illimination with compound image transforms. In Y. Zhang (Ed.), *Advances in face image analysis - techniques and technologies* (pp. 301–315). IGI Global Pub.

Shamir, L. (2012). Computer analysis reveals similarities between the artistic styles of van gogh and pollock. *Leonardo*, *2*(45), 149–154.

Shamir, L., Delaney, J., Orlov, N., Eckley, D., & Goldberg, I. (2010). Pattern recognition software and techniques for biological image analysis. *PLoS Computational Biology*, *11*(6), 1–10.

Shamir, L., Orlov, N., Eckley, D., Macura, T., Johnston, J., & Goldberg, I. (2008). Wndchrm - an open source utility for biological image analysis. *Source Code for Biology and Medicine*, (3), 1–13.

Shamir, L. & Tarakhovsky, J. A. (2012). Computer analysis of art. *ACM Journal on Computing and Cultural Heritage*, *2*(5).

Shepard, A. (1961). *Ceramics for the archaeologists*. Washington: Carnegie Institution Publication.

Sinopoli, C. (1991). *Approaches to archaeological ceramics*. New York: Plenum Press.

Spaulding, A. C. (1953). Statistical techniques for the discovery of artifact types. *American Antiquity*, (18), 305–313.

Tamura, H., Mori, S., & Yamavaki, T. (1978). Textural features corresponding to visual perception. *IEEE Trans on Systems, Man and Cybernetics*, (8), 460–472.

Teague, M. (1980). Image analysis via the general theory of moments. *Journal of Optical Society of America*, (70), 920–930.

Theobald, D. (2017). *Machine learning for absolute beginners*. Scatterplot Press.

Thér, R. (2015). Identification of pottery-forming techniques using quantitative analyis of the orientation of inclusions and voids in thin sections. *Archaeometry*, (58), 222–238.