Project Report

Multi-mineral classification of Apollo 15 Lunar microscopy samples through machine learning.

David Bloomer
MSc Data Science
2018-2020

Supervisor: David Weston
Birkbeck College, University of London

Table of Contents

1. Abstract

This study analyses the potential to perform multi-mineral classification of Lunar basalts through machine learning, using microscopy data collected from Apollo 15 mission samples. Sample mineralogy is predominantly composed of Pyroxene, Plagioclase Feldspar and fine-grained matrix. Labels are provided through scanning electron microscope derived QEMSCAN mineral interpretations. Innovative methodologies are applied to standardise image quality through pre-processing and perform pixelwise alignment between microscopy images and QEMSCAN data creating a dataset of 66.3M pixels. A wide range of features are considered to represent variation in mineral colour, structural and textural response of the minerals captured through imaging. Following analysis of pre-processing and model classification pipelines, a gradient boosting machine classifier is defined using an optimised feature set and parameterisation. Model performance is strongly driven by colour features, showing an overall accuracy of 81.1% and Dice similarity coefficient of 0.66. Performance is predominantly achieved through accurate prediction of predominant mineralogies, struggling to correctly classify Olivine and Silica, for which feature properties are demonstrated to overlap significantly with Pyroxene. Finally, an end-user web-based application is created to demonstrate potential usage of the classification model for use in routine analysis of rock composition.

2. Introduction

Routine analysis of rock composition is performed within planetary science as a preliminary step towards understanding of sample formation and evolution. This study aims to investigate the potential to perform high-level rock classification of Lunar basalts through machine learning, using microscopy data collected from Apollo 15 mission samples. In order to achieve this goal, limitations in the uniqueness of mineral properties and the digital representation of samples must be overcome. Automatisation of multi-mineral classification has implications in several domains where knowledge of rock composition or behaviour is vital.

2.1 Background

Computer vision has been applied in multiple domains where tasks involving use of digital images or video performed by humans are automated to improve efficiency and quality of analysis. Classification of rock composition and texture through image data is one such application.

Knowledge of rock composition is important in domains including construction, engineering, energy and planetary science, providing insights towards understanding physical and chemical behaviour, formational setting, evolution, and use in calibration to other data sources.

Literature contains examples of multiple machine learning approaches to analyse specific aspects of a rock's composition and texture, to determine grain-size, rock classification, binary (grain-porosity) or multi-mineral classification. Multi-mineral classification is the most challenging of these problems, where semantic segmentation is performed to classify minerals at the pixel level.

Rock forming minerals possess highly variable optical and morphological properties, in response to changes in phase and structure (Nesse, 2004) (Figure 2.1). In polarized-light microscopy, thin sections prepared from a rock may be viewed by a human interpreter under plain (PPL) and crossed polarized (XPL) light sources. Interpretation of minerals requires the identification of multiple overlapping properties, which reflect changes in structure, texture and refractive characteristics such as the relationship between colour and polarised light orientation (pleochroism). Once minerals are classified, relative proportions are calculated through box counting, allowing for rock classification. Textural observations such as grain size, shape and distribution are produced by enlarging sections through projection. While these skills require years of training, interpretation remains highly time consuming, subjective and lacks repeatability.

Rock classification through digital microscopy is more challenging still, as data are commonly collected at a single angle, limiting the potential for mineral identification through pleochroism. Digital data will also possess non-physical variation in optical properties, sourced through lack of consistency in photographic equipment and technique. Standardised approaches to data collection are proposed (Berrezuetta *et al*, 2019), but are not common practise.

Laboratory based methods are available that provide alternate data sources for use in rock imaging and classification of composition, the most common of which are x-ray diffraction (XRD), scanning electron microscope (SEM) and x-ray computed microtomography (µCT) based techniques. These are vital for the analysis solid solutions or fine-grained materials not suitable for microscopy (Raith *et al*. 2012). However, each have disadvantages in terms of cost, analysis time, lack of textural classification, and are therefore used sparingly in combination with microscopy.

2.2 Project Objectives

This project aims to investigate the potential to design an interactive solution to perform high-level rock classification of Lunar basalts, as specified through the following objectives:

- Establish an image pre-processing workflow suitable for aligning data from sources with varying rotation, scale, translation and geometric distortion.
- Quantify the importance of colour, structural and textural features in Lunar basalt mineral classification.
- Evaluate the suitability of supervised machine learning approaches based on an assessment of computational efficiency and performance metrics.
- Demonstrate the potential for an end-user web-based application focusing on rock classification.

A thorough literature review has been performed and is used to guide the choice of features and models explored (see Project Proposal). To the best knowledge of the author, multi-mineral rock classification has not previously been attempted using the mineral suite characteristic of Lunar basalts. Within the broader context of machine learning, several approaches to multi-mineral classification are documented, with literature citing test accuracies of up to 93.81% (Izadi *et al*., 2017).

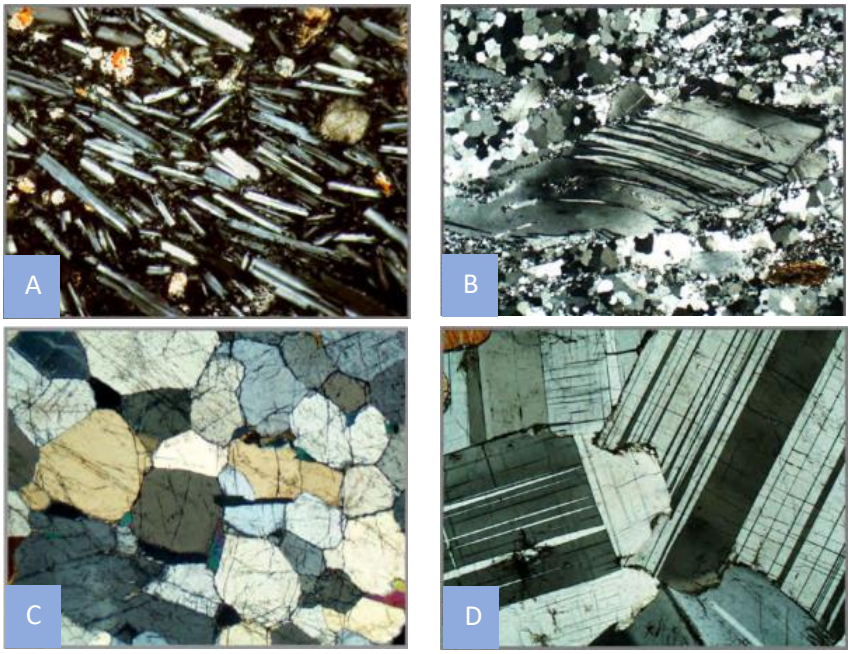


Figure 2.1: Example of single mineral, Feldspar, represented within plain polarized-light microscopy images, demonstrating variability in colour and texture: A. Basalt; B. Deformation and twinning; C. Recrystallisation; D. Polysynthetic twinning. Raith *et al*. (2012).

2.3 Data

The dataset comprises polarized-light microscopy and quantitative mineralogy digital images from four Apollo 15 basalt samples. Sample mineralogy is typical of Lunar basalts, being predominantly composed of Pyroxene, Plagioclase Feldspar and Olivine. Plain (PPL) and crossed polarized-light (XPL) microscopy images are provided by the School of Earth and Environmental Sciences at the University of Manchester. Quantitative mineralogy (QEMSCAN) images are released under Creative Commons 4.0 licensing through Mendeley Data. Details of the data collection methodology are outlined in Bell *et al*. (2020). A summary of mineral classes, proportions and image properties are given in Table 2.1.

A dataset containing scanning electron microscope (SEM) derived quantitative mineralogy interpretation data has been chosen for use in sample validation over manual segmentation to minimise interpretation bias.

PPL and XPL microscopy images are available as overlapping tiles at 3.06 microns per pixel resolution, with QEMSCAN data as full sample images at 5 microns per pixel resolution (Figure 2.2).

Sample	**Split**	**Mineral Classes and Proportions (>1%)**	**Image Properties** *n, pixels (format)*
15125	6	Matrix (51.4%), Pyroxene (45.3%), Olivine (3.1%)	PPL/XPL: 4, 1920x2560 (tif) QEMSCAN: 1, 8035x7266 (tif)
15475	15	Pyroxene (59.7%), Plagioclase Feldspar (30.2%)	PPL: 10, 1920x2560 (tif) XPL: 18, 1920x2560 (tif) QEMSCAN: 1, 9463x9262 (tif)
15555	209	Pyroxene (52.4%), Plagioclase Feldspar (30.4%), Olivine (12.1%)	PPL/XPL: 12, 1920x2560 (tif) QEMSCAN: 1, 11958x14256 (tif)
15597	18	Pyroxene (59%), Matrix (35%)	PPL/XPL: 4, 1920x2560 (tif) QEMSCAN: 1, 8218x7012 (tif)

Table 2.1: Technical summary of Apollo 15 basalt sample dataset comprising plain (PPL), crossed polarized-light (XPL) microscopy and quantitative mineralogy (QEMSCAN) images. Mineral proportions after Grove and Walker (1977), Schnare *et al*. (2008), Longhi *et al*. (1972), Weigand and Hollister (1973).

Note: *Section 2.3, Table 2.1, Figure 2.1 and 2.2 are from the Project Proposal.*

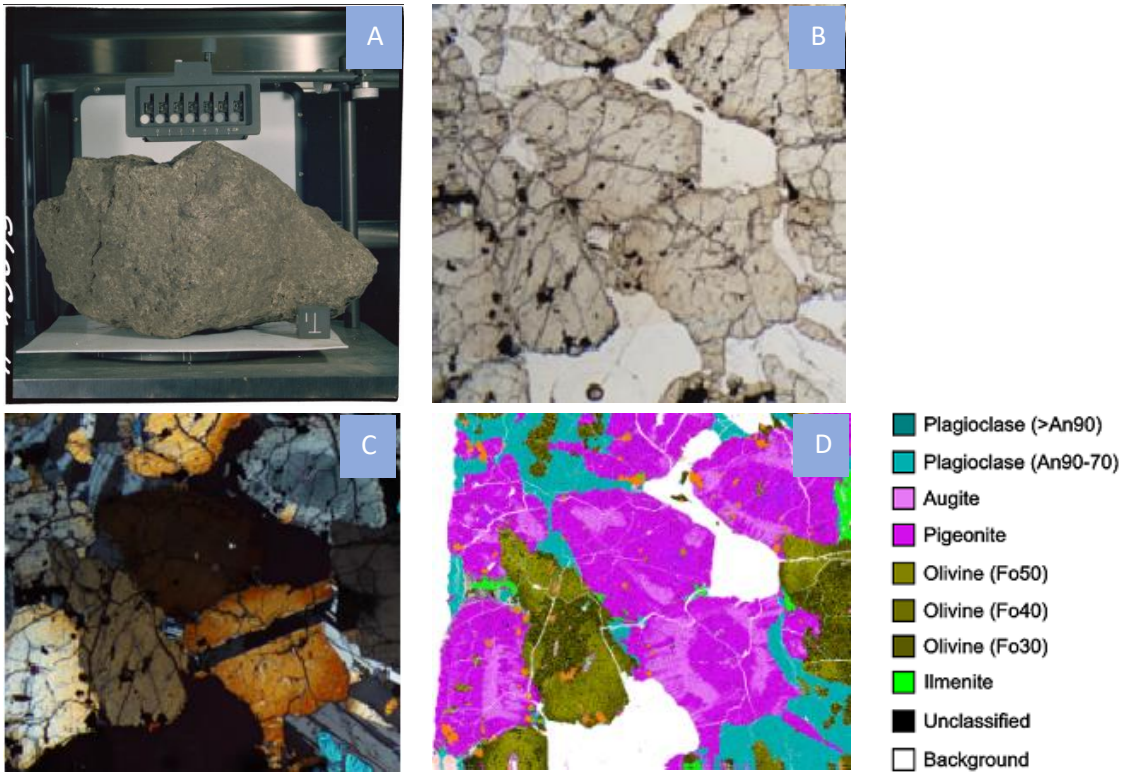


Figure 2.2: Apollo 15 Sample 15555 ("Great Scott") images, Split 209: A. Laboratory sample (LPI Lunar Sample Atlas, 2020); B. Plain polarized-light microscopy (5x magnification); C. Cross polarized-light microscopy (5x magnification); D. Quantitative mineralogy (QEMSCAN) (Bell *et al*., 2020). Note: Augite and Pigeonite are phases of the Pyroxene mineral group.

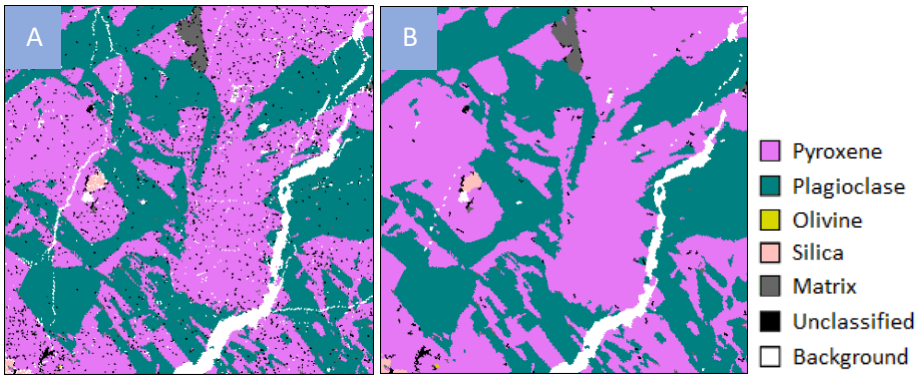


Figure 3.1: QEMSCAN interpretation (Sample 15475, Split 15): A. Reduced mineral interpretation; B. Reduced resolution ('denoised') interpretation.

3. Method

A method is outlined to address the project goals. Initially data are prepared for ingestion within a machine learning model through class definition, improvement of image quality, pixelwise alignment of the dataset with interpretation labels and feature engineering to characterise variation in colour, shape and texture. Model selection is performed using a custom-built pipeline, with optimal model parameterisation considered through feature selection and hyperparameter tuning.

The code used to perform the method is provided in a separate digital repository. Details of the repository directory structure are provided in Appendix 9.3.

3.1 Image Pre-processing

Image pre-processing aims to resolve non-physical variation in image properties, which manifest both through difference in resolution between PPL, XPL and QEMSCAN data, and through capture of microscopy digital images.

QEMSCAN images provide quantitative interpretation of mineral phase represented by RGB colours. The RGB classification scheme is mapped to a sequential integer representation through use of a python dictionary, with mineral phases combined to provide classification at the mineral rather than phase level (Figure 3.1A). Defined classes are Pyroxene, Plagioclase Feldspar, Olivine, Silica, matrix, unclassified and background. Where appropriate, mineral groups have been combined based on understanding of data limitations, for example, combining accessory phases (Ilmenite and Chromite) and fine-grained minerals under the classification matrix, due to their inability to transmit light.

Interpretation data possesses significant small-scale variability in mineralogy associated with fine-grained textures, and potentially instrument sourced random noise. It is desirable to remove this variability, given it exists at a higher resolution that would be possible to predict from the microscopy data and is later seen to introduce difficulties in estimating features used for alignment (Section 3.2).

To better align the effective resolution of both images, the python opencv library is used to calculate connected components within QEMSCAN images using Fast Block Based Labelling (Grana *et al*., 2009). Components below a threshold of 6 pixels in size are reassigned to the background classification. Morphological closing is then applied to the background label, using dilation and erosion operators with a 2x2 followed by 3x3 kernel to minimise operator artefacts. The outcome is that intra-mineral gaps are filled by the appropriate surrounding label(s), producing a cleaner image at reduced resolution as demonstrated in Figure 3.1.

A further potential mechanism to deal with differences in resolution is introduced through convolution of microscopy data with a 3x3 Gaussian filter kernel, the influence of which is assessed as a pre-processing step within model selection (Section 3.4).

Two contrasting approaches are commonly used when dealing with variation in image capture. Image equalisation aims to standardise the digital representation of data prior to learning. Several techniques within the python scikit-image library were considered in the RGB and HSV colour spaces, based on histogram manipulation, rescaling and gamma adjustment. The most promising technique is Contrast Limited Adaptive Histogram Equalization (CLAHE) (Figure 3.2), its influence on model performance further considered as pre-processing step within model selection (Section 3.4).

An alternate approach to equalisation is data augmentation, where it is considered that introducing additional artificial variability produces a more robust classification. Data augmentation is commonly applied within deep learning workflows and has been shown to increase the potential for model generalisation (Wong *et al.*, 2016). A data augmentation pipeline is created using the python tensorflow library to demonstrate simulated variation in texture orientation (flip and rotation), imaging (brightness, saturation and contrast), image resolution (crop and scaling) and image quality (Gaussian noise) (Figure 3.3). The impact of data augmentation is not be further considered due to the lack of inclusion of deep learning approaches within the revised scope of this project.

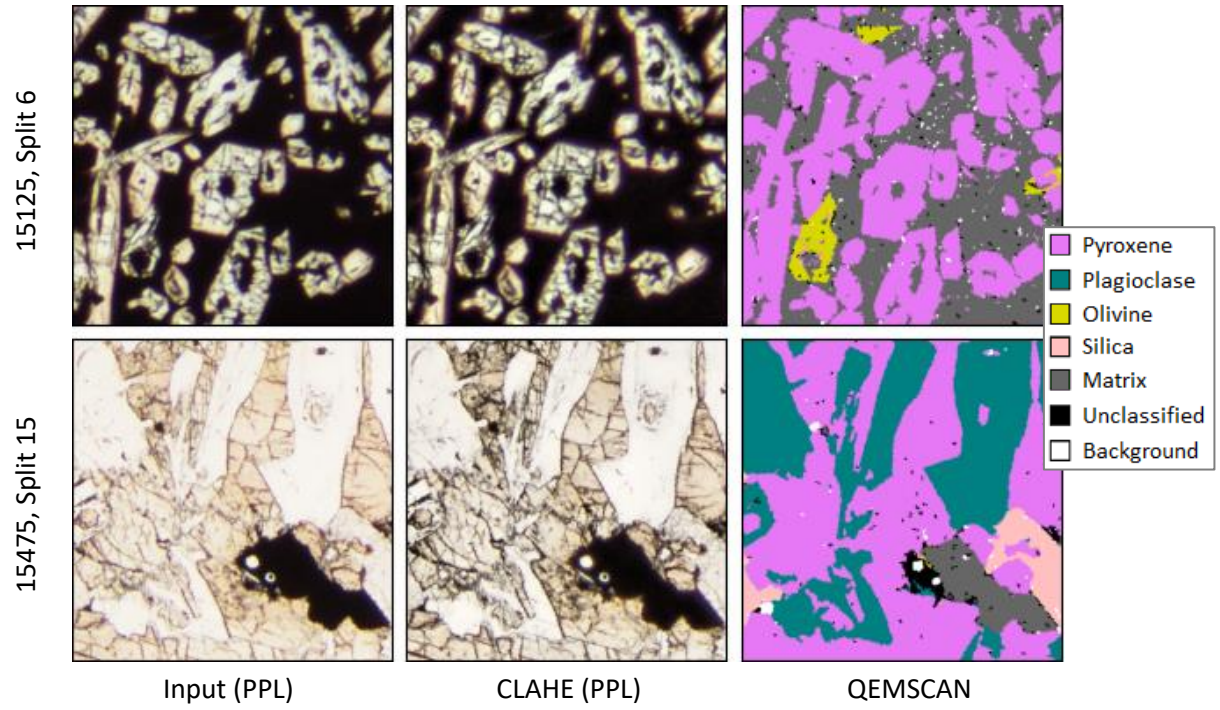


Figure 3.2: Image standardisation (Top: Sample 15125, Split 6, Bottom: Sample 15475, Split 15,): Left. Input PPL; Center. Contrast Limited Adaptive Histogram Equalization (CLAHE); Right. QEMSCAN interpretation. Expanded view containing all samples and XPL is available in **Appendix 9.1**.

3.2 Image Alignment

To prepare data for machine learning ingestion, spatial alignment between digital microscopy and mineral interpretation images is required, as performed through pixelwise registration between PPL, XPL and QEMSCAN data.

The dataset provided several challenges. QEMSCAN data are available full sample images, while microscopy data are available as overlapping tiled images. Microscopy images possess negative geometrical distortion around the image edge, typical of images source from stereoscopic microscopes. Finally, PPL and XPL images do not possess a common sampling strategy and are not always aligned or available in equal proportions (Table 2.1).

An initial approach was considered to stitch PPL and XPL microscopy images prior to alignment to the QEMSCAN interpretation. However, it was found that the Brown and Lowe (2007) image stitching approach suffered from a lack of repeatability, due to variation in feature matching introduced by image orientation, zones of narrow overlap, geometrical distortion, variation in brightness levels, mineral colour and texture through sample orientation and background interference in image blending. Furthermore, the process resamples images at a reduced resolution, creating the potential for data loss.

To mitigate for these factors, a two-stage workflow was devised using functionality contained within the python opencv library to iteratively align the QEMSCAN interpretation to microscopy images, keeping variation in colour and texture introduced through orientation, whilst ensuring no data loss occurred:

1. Affine-rigid homography matrix, to compensate for rotation, scale and translation.
2. Affine homography matrix, to compensate for small-scale negative geometrical distortion.

In each step, features are estimated from either the PPL or XPL greyscale microscopy image and a simplified high contrast mineralogical (pyroxene only) QEMSCAN interpretation, following removal of connected components under a threshold of 40 pixels, as described in Section 3.1. Features are detected through Scale-Invariant Feature Transform (SIFT) (Lowe, 1999) (Figure 3.4), as the open source ORB algorithm proved ineffective in resolving rotation. Fast Library for Approximate Nearest Neighbors (FLANN) (Muja and Lowe, 2012) is used to match features, which are then refined using Lowe's ratio test (Lowe, 1999).

Following this process, the refined feature set is used to a create homography matrix, which is applied to perform an affine-rigid then affine warp to the QEMSCAN interpretation iteratively. Finally, an alignment is performed between PPL and XPL microscopy images where required, and outputs cropped to remove rows or columns entirely composed of the background to increase storage and computational efficiency.

Alignment was performed for 21 of 40 available microscopy images, creating a dataset of 66.3M pixels. An example of the alignment between images is shown in Figure 3.5. Alignment of the remaining images was not possible, despite numerous attempts at various pre-processing or alignment strategies, including image pre-processing or conversion to gradient prior to alignment, brute-force matching, confidence-based match filtering and use of projective homography matrices.

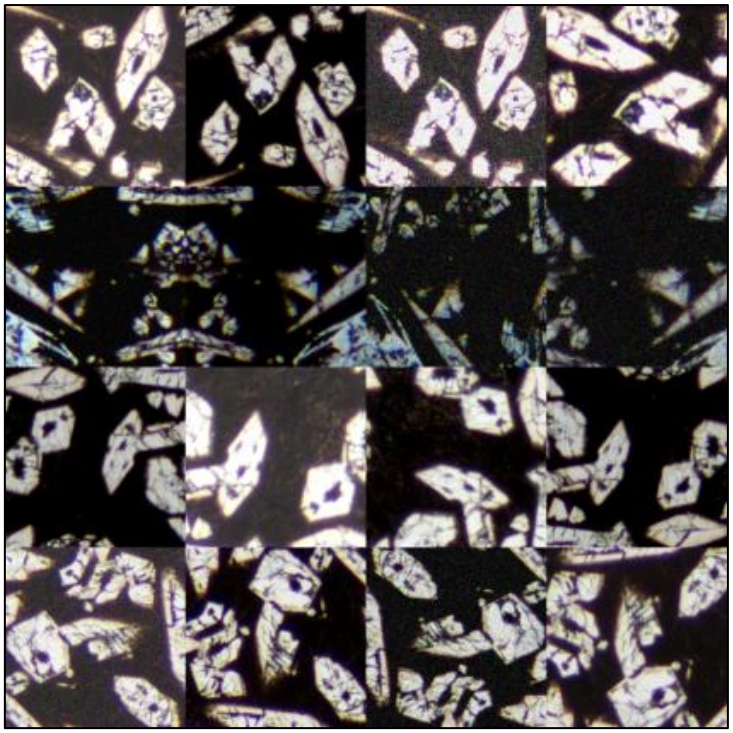


Figure 3.3: Data augmentation (Sample 15125, Split 6): Rows represent different sub-samples from PPL image, columns represent randomised augmentation simulations.

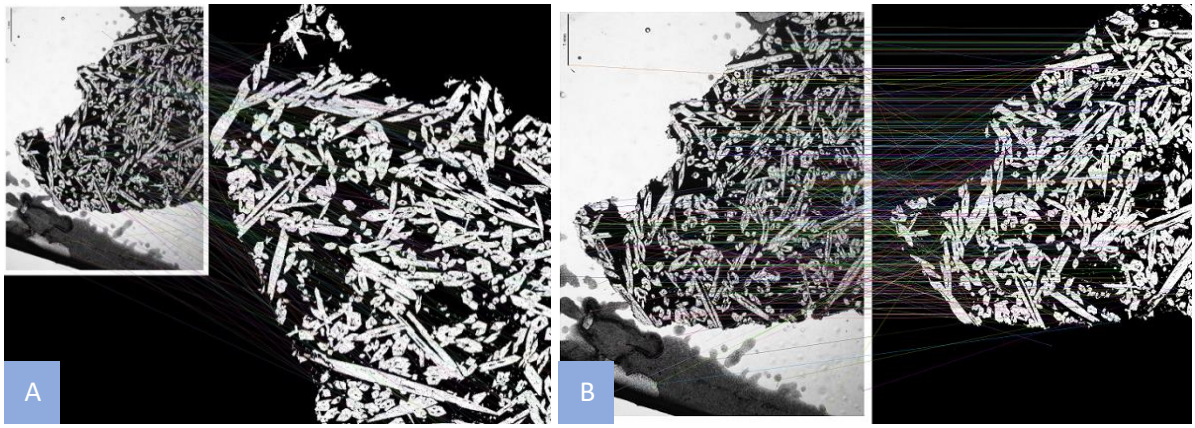


Figure 3.4: Feature matching (Sample 15125, Split 6): Features detected using SIFT are shown between the greyscale PPL microscopy image and simplified mineralogy QEMSCAN interpretation: A. Stage one (affine-rigid); B. Stage two (affine).

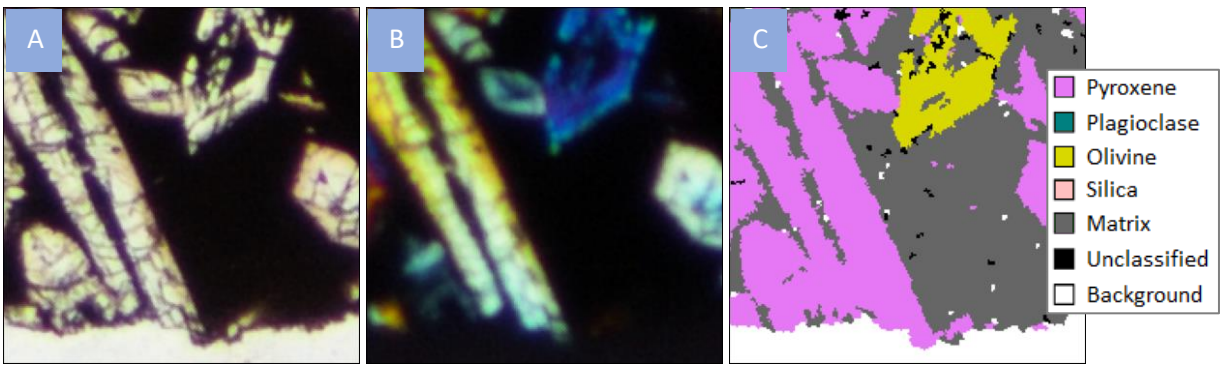


Figure 3.5: Aligned images (Sample 15125, Split 6). A. PPL; B. XPL; C. QEMSCAN interpretation.

3.3 Feature Engineering

A combined 114 features are generated from aligned PPL and XPL microscopy images. A range of features are considered for use in discriminating variation in colour, shape and texture. Unless otherwise stated, features were created using the python scikit-image library. Created features are broadly classified into colour, edge, ridge, filter bank and neighbourhood categories.

3.3.1 Colour Representation

Colour is a key attribute for interpretation of minerals, despite its high degree of variability with respect to angle of orientation (Figure 2.1). Greyscale intensity, HSV and CIELAB have been shown to have a positive influence in multi-mineral classification over the RGB colour model (Dunlop, 2006; Maitre *et al*., 1999; Thompson *et al*., 2001; Rubo *et al*, 2019). HSV and CIELAB both represent colour relative to wavelength, better approximating human colour perception (Fairchild, 2003), overcoming limitations in the sensitivity of the RGB colour model to variation in light source.

As an alternative to image equalisation through pre-processing (Section 3.1), an additional set of features are created through z-normalisation of each transformed colour channel (Equation 3.1). This approach follows an assumption that image standardisation has the biggest impact on quantitative colour representation, whereas textural features highlight relative differences in image properties.

$$Z_n = \frac{(n - \bar{Z})}{\sigma_Z}$$

Equation 3.1: Z-normalised colour representation, as applied to each colour channel for greyscale, RGB, HSV and CIELAB colour representations.

3.3.2 Edge Filters

Canny and Sobel filters are considered to define or emphasise edges as estimated through use of the image gradient: a directional change an images colour or intensity. The Canny filter uses a Gaussian derivate, while Sobel uses a pre-defined operator. Canny performs hysteresis thresholding to produce a binary interpretation of edges. Use of edge filters may be relevant for high contrast minerals that are limited in lateral extent, such as accessory phases. Use of a Canny filter with a Gaussian standard deviation of 1 is seen to give a comparable impression of mineral edges to Sobel, while Sobel retains its approximation of edge intensity (Figure 3.6).

3.3.3 Ridge Filters

An extension of edge detection within image gradient, ridge filters use eigenvalues of a Hessian matrix composed from second-order derivatives of image intensity to detect ridge structures the image gradient. Hessian, Frangi, Sato and Meijering ridge filters are considered to highlight intra-mineral variation, such as that associated with structure (mineral cleavage) (Figure 3.7).

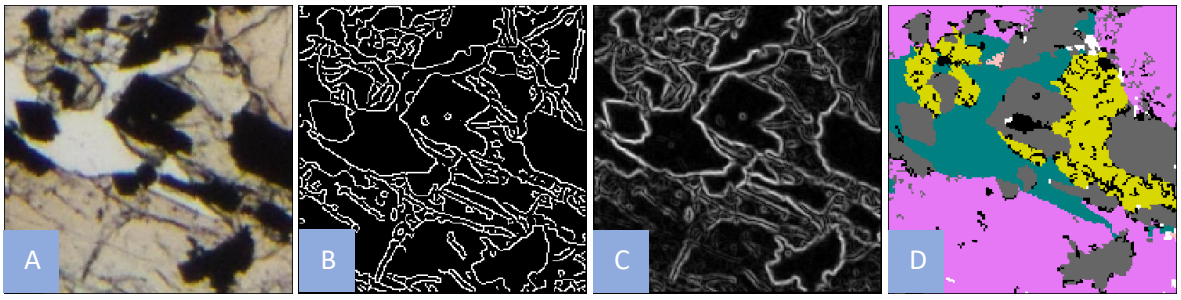


Figure 3.6: Edge filters (Sample 15555, Split 209): A. PPL; B. Canny (σ=1); C. Sobel; D. QEMSCAN interpretation (see **Figure 3.5** for key).

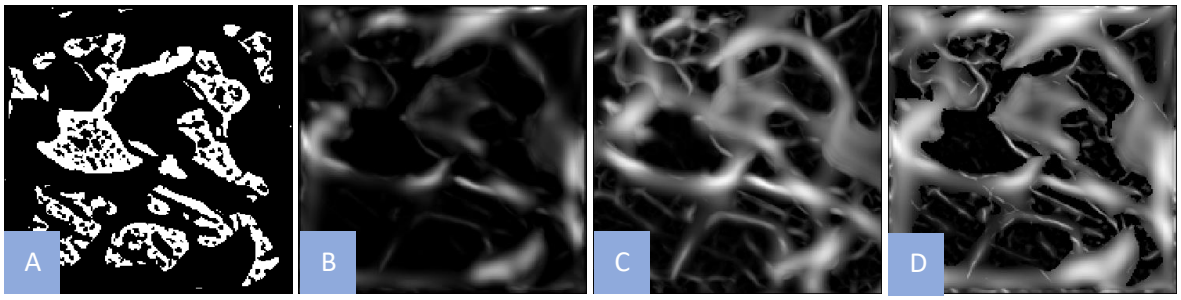


Figure 3.7: Ridge filters (Sample 15555, Split 209): A. Hessian; B. Frangi; C. Sato; D. Meijering. Input image and mineral interpretation are consistent with **Figure 3.6**.

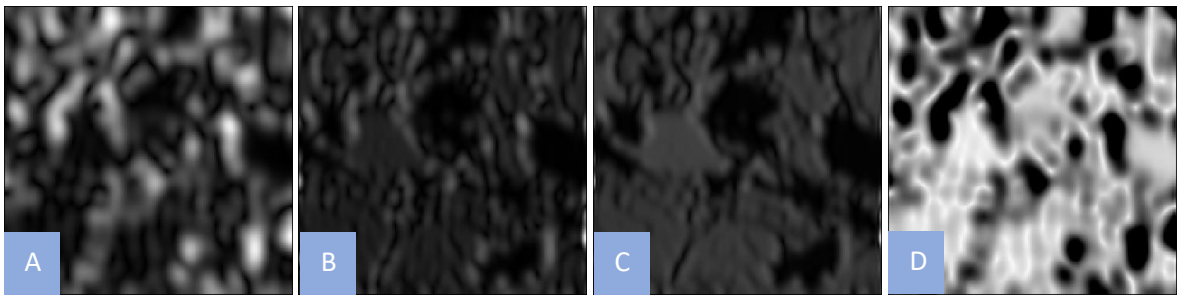


Figure 3.8: Gabor filter bank (Sample 15555, Split 209): A. 0 radians (θ), 0.1 frequency (ω); B. 0θ, 0.2ω; C. 0θ, 0.3ω; 4. Gabor frequency gradient. Input image and mineral interpretation are consistent with **Figure 3.6**.

3.3.4 Filter Banks

Filter banks provide a systematic approach to defining linear filters that respond to directional and frequency characteristics of an image and are considered to for their extensive use in computer vision to represent texture. The Gabor filter bank performs a Fourier transform through application of Gaussian functions in the spatial and frequency domains (Clark *et al.*, 1997), and is defined using frequencies of 0.1, 0.2, 0.3 and 0.4, and orientations of 0/π4, 1/π4, 2/π4 and 3/π4 radians (Figure 3.8A-C).

An extension of the Gabor filter bank is devised following the observation of Jiang *et al*. (2018) that specific minerals demonstrate characteristic distributions of Gabor filter responses (Figure 3.9). At a constant angle of 0 radians, a least squares regression is fit to Gabor filter bank frequencies expressed as a linear system of equations and defined as the Gabor frequency gradient (Figure 3.8D). As this feature is derived from products of the Gabor filter bank, it is computationally relatively inexpensive.

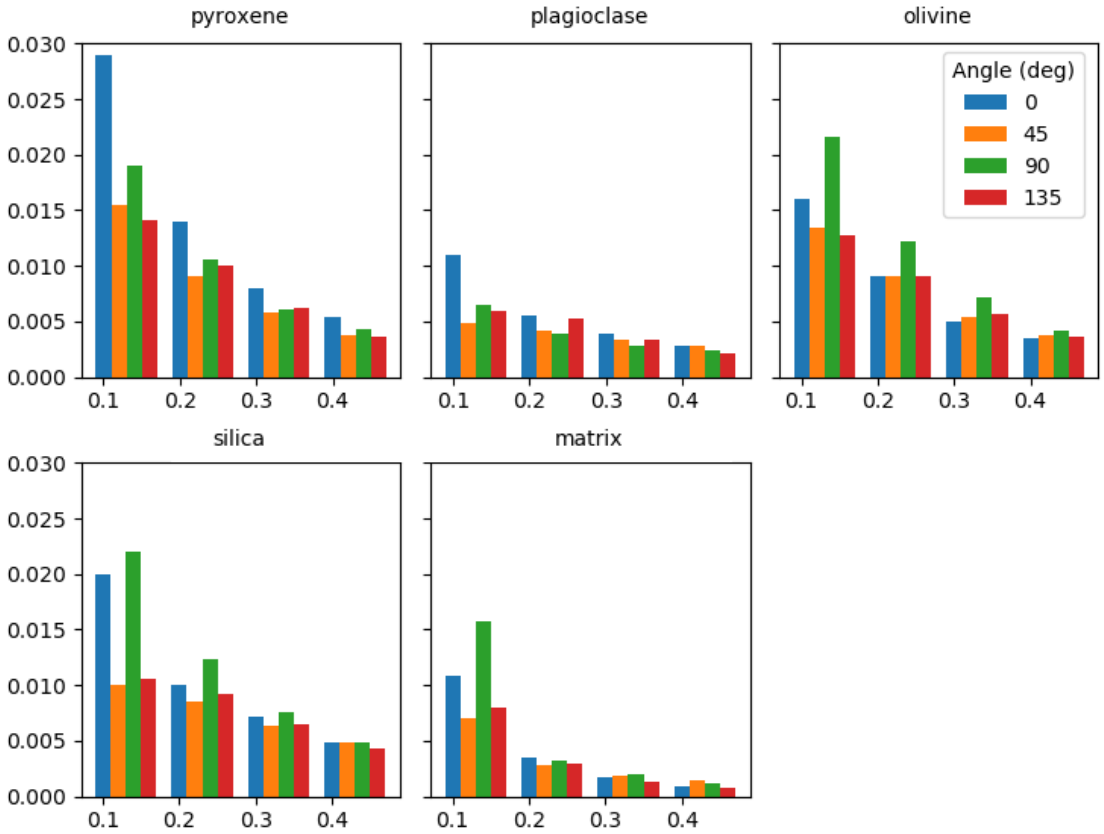


Figure 3.9: Gabor filter bank (from all samples): distribution of Gabor frequency and orientation filters by mineral class. Pyroxene, Olivine and Silica are seen to show comparable profiles.

3.3.5 Neighbourhood Features

Neighbourhood features are defined as those involving a comparison between a pixel and its neighbouring pixels, encompassing both Haralick and Local Binary Pattern features. Neighbourhood features are extensively used for their potential to represent textures and repeating patterns within computer vision.

Haralick features are derived from a gray level co-occurrence matrix (GLCM) (Haralick *et al*., 1973) describing offset patterns between pixels of similar intensity. Haralick features calculated are contrast, dissimilarity, homogeneity, angular second moment, energy, correlation (Figure 3.10). Implementation of Haralick feature extraction is taken from Jónathan Heras (2017). Computation of the GLCM is required at each pixel within an image, describing the distribution of intensity within a sliding window of 11x11 pixels.

Local Binary Pattern (LBP) features are derived by determining the relative intensity of a pixel to its neighbours at a specified radius, with the distribution of counts characterising the surrounding texture, as popularised by Ojala *et al*. (2002). An extension of the LBP is the uniform pattern, which provides a rotation invariant descriptor. Non-uniform and uniform LBPs are calculated at pixel radii of 2,4,8 and 16 (Figure 3.11).

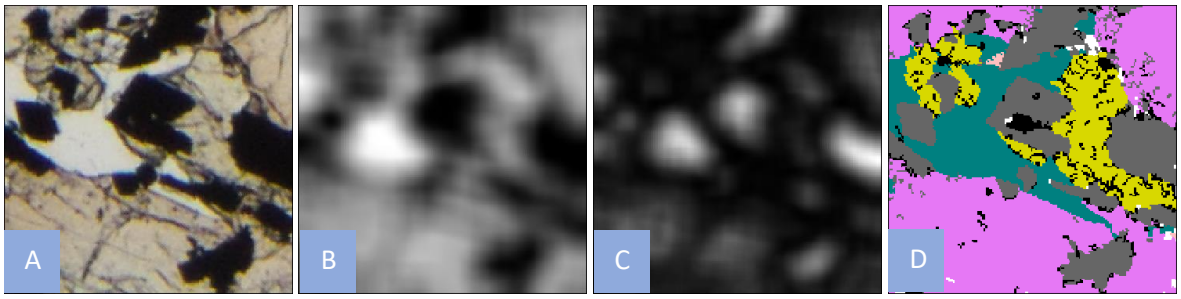


Figure 3.10: Haralick features (Sample 15555, Split 209): A. PPL; B. Contrast; C. Energy; D. QEMSCAN interpretation (see **Figure 3.5** for key).

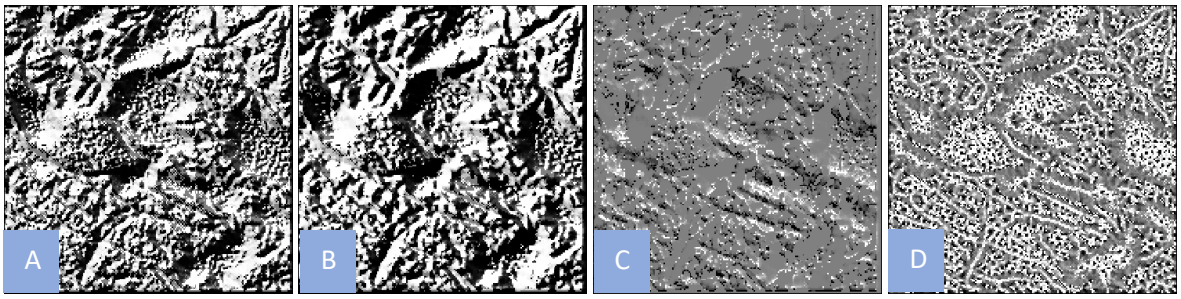


Figure 3.11: Linear Binary Pattern (LBP) features (Sample 15555, Split 209): A. LBP Radius 2; B. LBP Radius 3; C. LBP Radius 4; D. Uniform LBP Radius 2. Input image and mineral interpretation are consistent with **Figure 3.10**.

3.3.6 Feature Computation

Given the objective of the project is to create an interactive application driven solution, computational efficiency of calculating features is of key importance. Figure 3.12 demonstrates the relationship between image pixel count and computation efficiency expressed through calculation time. While all features follow O(n) complexity, Haralick features demonstrate a considerable overhead in calculating the GLCM per pixel and are therefore excluded for consideration in model training. However, their impact for single attribute analysis is further considered in Section 4.2.

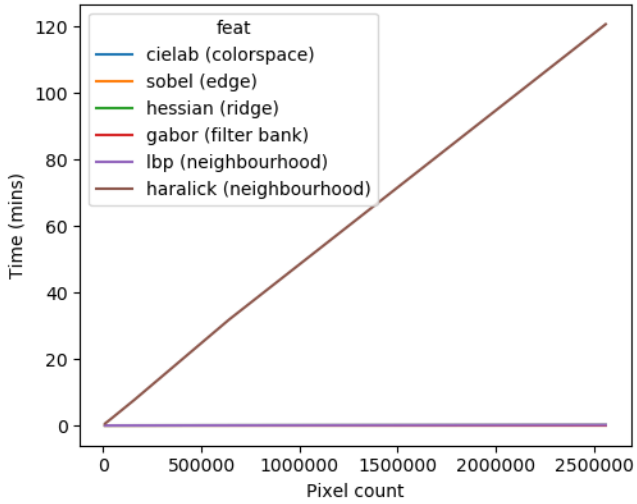


Figure 3.12: Feature creation computational efficiency. A single feature is shown for each feature class, additional features within each class showed comparable computation times.

3.4 Feature and Model Evaluation

To optimise the efficiency of the model evaluation process, from the datasets 66.3M samples and 114 features, a reduced dataset of 15.3M samples is selected for use in a custom-built pipeline to systematically assess image pre-processing and model selection. The reduced dataset is based on a single image from each rock sample, capturing a wide variability in mineral response. Data corresponding the unclassified and background classification label are excluded from training. A test-train-split strategy is used to determine test performance, with a stratified sample of 30% taken to create a test dataset. The pipeline is built using the python scikit-learn library API.

As the dataset is imbalanced, reflecting the mineral proportions of the underlying data (Table 2.1), two approaches are considered to compensate for class imbalance: training models using an under - sampled balanced dataset or use of ensemble models supporting class weights or under-sampling strategies. It was found that model performance of the two approaches is comparable, and so the impact of class imbalance is assessed through model selection. Further consideration is given to class imbalance through selection of performance metrics. While literature for multi-mineral classification commonly reports model performance through accuracy and weighted accuracy, model performance will additionally be assessed using the Dice (F1) similarity coefficient. The Dice coefficient is equivalent to intersection over average pixel count, and so provides a more accurate representation of model performance where class imbalance is present.

The range of classification models has a heavy bias towards tree-based ensemble methods, which have been shown within literature to provide strong performance, reasonable training times and flexibility to deal with large imbalanced datasets. Classification models within the scikit-learn library are supplemented through additional models within the imblearn and lightgbm libraries. The following classification models are evaluated:

- Linear Support Vector Machine with Stochastic Gradient Decent (LSVM)
- Random Forest (RF)
- Random Forest with Balanced Class Weights (RF[W])
- Balanced Random Forest (BRF)
- Gradient Boosting (GB)
- LightGBM (LGBM))
- LightGBM with Balanced Class Weights (LGBM[W])

Evaluation training times for each classification model are shown in Figure 3.13, with the exception of GB, which failed to train within 24 hours, in part due to the lack support for parallelisation within the scikit-learn implementation. Model performance for each image pre-processing scenario are given in Table 3.1, and relative metric performance associated with no pre-processing visualised in Figure 3.14.

Application of a Gaussian filter to PPL and XPL inputs results in an average Dice performance increase of 4.7% for all models, although is notably worse for LSVM. Adaptive equalisation (CLAHE) and colour z-normalisation have minimal impact on model performance, likely due to all images being sampled from a common piece of equipment. However, both equalisation techniques are sensitive to image cropping, where image extend has a strong impact on absolute colour representation, and therefore on predictive performance (Figure 3.15). This is of key concern given the goal of developing an application, where an end-user has control of the image sampling.

While RF based techniques gave both the highest accuracy and Dice performance metrics, the model was suspected of over-fitting based on performance drop-off where maximum tree length is reduced (Table 3.1). A comparison of RF, LGBM and LGBM[W] model predictions on out of sample images (Figure 3.16) show that RF and LGBM possess similar performance, failing to accurately classify Olivine and Silica.

By contrast, in attempting to improve classification of Olivine and Silica which possess lower sample proportion, LGBM[W] (and the other models compensating for class imbalance) fail to accurately capture the distribution of Pyroxene (Figure 3.16). It is theorised that the properties of the features representing Olivine, Silica and Pyroxene overlap, limiting the potential for their discrimination, which is further discussed in Section 4.2.

As a compromise between overall model performance, training time and sensitivity to image selection, LGBM is chosen to perform feature evaluation in combination with a Gaussian filter pre-processing workflow.

	Pre-Processing Workflows											
	None			Gaussian			Gaussian + CLAHE			Z-Normalisation		
Model	Acc	Acc[W]	Dice	Acc	Acc[W]	Dice	Acc	Acc[W]	Dice	Acc	Acc[W]	Dice
LSVM	0.766	0.472	0.771	0.770	0.477	0.466	0.764	0.467	0.462	0.782	0.494	0.477
RF	0.861	0.702	0.767	0.938	0.912	0.935	0.936	0.904	0.930	0.874	0.745	0.802
RF[W]	0.858	0.699	0.764	0.939	0.916	0.938	0.936	0.911	0.934	0.871	0.739	0.798
RF15*	0.809	0.548	0.582	0.828	0.608	0.662	0.820	0.579	0.629	0.818	0.585	0.628
RF[W]15*	0.731	0.765	0.629	0.775	0.827	0.690	0.764	0.826	0.677	0.731	0.792	0.638
BRF	0.764	0.817	0.675	0.839	0.898	0.796	0.830	0.894	0.779	0.765	0.836	0.679
LGBM	0.807	0.603	0.646	0.820	0.650	0.695	0.817	0.641	0.691	0.816	0.624	0.668
LGBM[W]	0.698	0.749	0.600	0.726	0.785	0.635	0.716	0.784	0.624	0.710	0.778	0.616

Table 3.1: Model comparison test errors, for various approaches for image standardisation. 15* represents that the default parameterisation of RF was overwritten to set max_length=15.

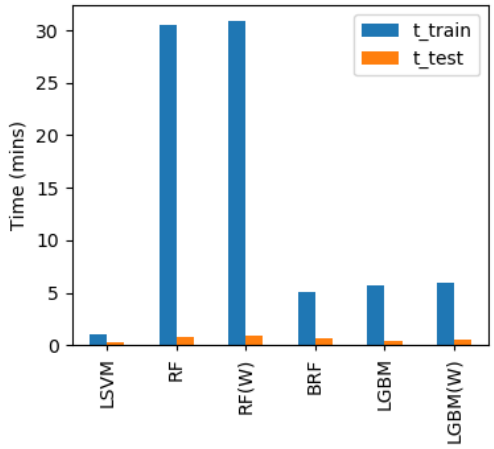


Figure 3.13: Training (t_train) and test (t_test) evaluation times in minutes for a single training run.

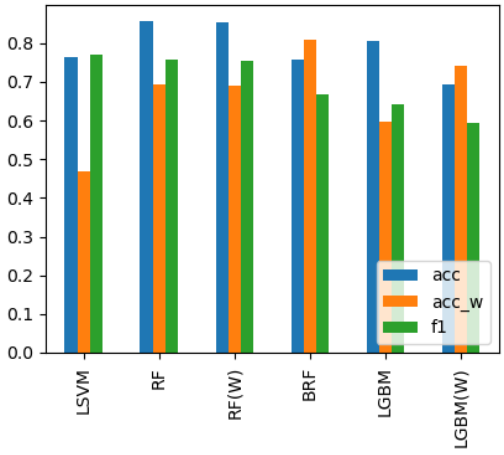


Figure 3.14: Estimated model performance through test evaluation. Random Forest (RF) model is interpreted to over-fit, with LightGBM (LGBM) model chosen.

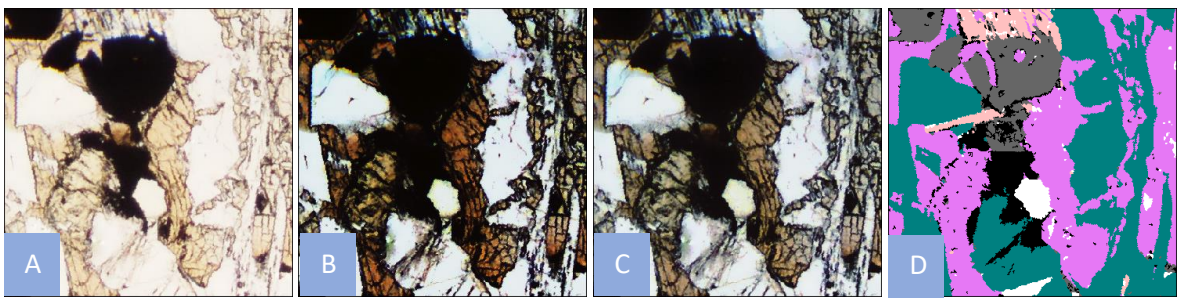


Figure 3.15: Influence of initial image extent on pre-processing, shown for Contrast Limited Adaptive Histogram Equalization (CLAHE) (Sample 15475, Split 15): A. PPL; B. Full sample CLAHE; Cropped sample CLAHE; D. QEMSCAN interpretation (see **Figure 3.16** for key).

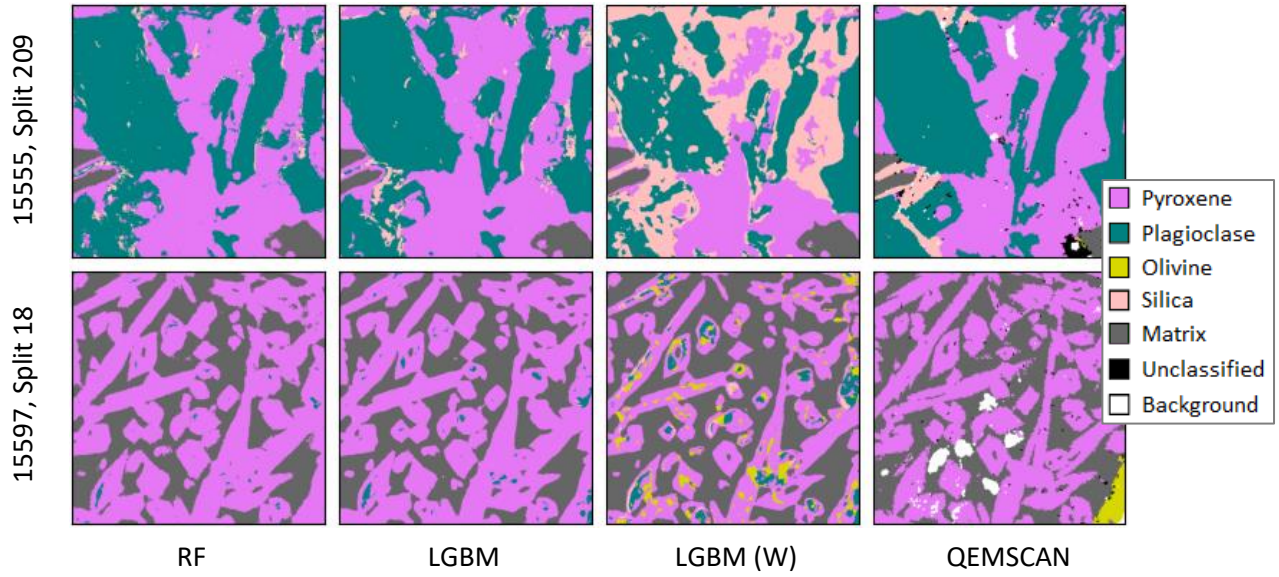


Figure 3.16: Model performance on out of sample images. Images for all samples in **Appendix 9.2**.

Recursive Feature Elimination (RFE) is performed to rank features in terms of relative importance. The top 16 features as estimated through RFE are shown in Table 3.2. Features within the colour representation group dominate, making up all the features except for Sato, Uniform Linear Binary Pattern with a radius of 4 pixels and a single component of the Gabor filter bank. An even distribution exists between features calculated from PPL and XPL inputs.

The relationship between model performance and the top n features estimated through RFE is shown in Figure 3.17. To decrease the potential for over-fitting, and reduce the computational overhead of the model, the top 30 features are selected for use within the final model.

Finally, hyperparameter tuning of the LGBM classifier is performed using a Bayesian optimization strategy as implemented through the python hyperopt library. The Bayesian objective function maximises the average Dice and weighted Dice similarity coefficient over 100 iterations but is seen to diverge on an optimal parameterisation much quicker (Figure 3.18).

Rank	Input	Feature	Component	Rank	Input	Feature	Component
1	XPL	CIELAB	B	9	PPL	HSV	S
2	XPL	RGB	R	10	PPL	Sobel	
3	XPL	CIELAB	A	11	PPL	Uniform LBP	Radius 4
4	PPL	RGB	R	12	XPL	Sato	
5	PPL	RGB	B	13	PPL	HSV	H
6	PPL	CIELAB	A	14	XPL	HSV	S
7	PPL	Sato		15	PPL	Gabor	0 θ, 0.1 ω
8	XPL	RGB	B	16	PPL	RGB	G

Table 3.2: Feature importance as estimated by Recursive Feature Elimination using the LightGBM Classification model and Gaussian filter pre-processing.

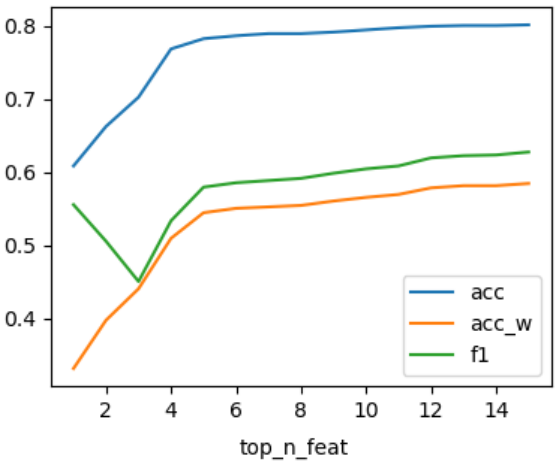


Figure 3.17: Model performance selecting the top n features selected by Recursive Feature Elimination, showing accuracy (acc), weighted accuracy (acc_w) and Dice (f1) similarity coefficient.

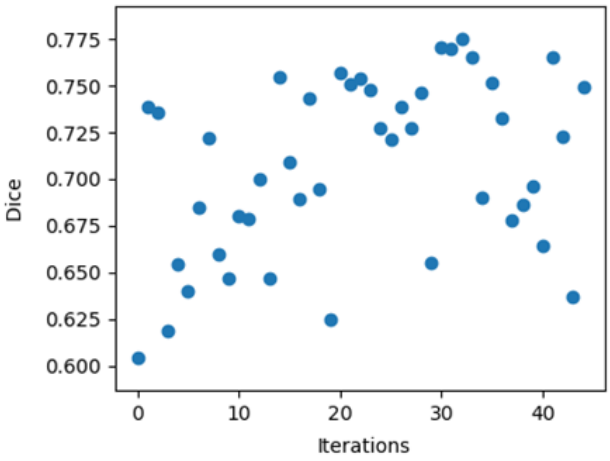


Figure 3.18: Bayesian hyperparameter optimisation, using LightGBM Classier, Gaussian filter pre-processing and reduced feature set. Optimal hyperparameters were estimated after approximately 30 of 100 iterations performed.

4. Evaluation and Discussion

The final model is trained using the full dataset of 66.3 pixels with optimal strategies for pre-processing, model selection, feature selection and parameterisation. Model performance is reported based on a 5 k-fold cross validation strategy. Furthermore, an attempt is made to understand model performance based on one-dimensional characterisation of the dataset.

4.1 Model Performance

Performance of the final model is summarised in Table 4.1. The model achieves an overall accuracy of 81.1% and Dice similarity coefficient of 0.66. Overall accuracy is achieved through strong predictive performance for predominant mineralogies within the dataset, while struggling to accurately classify Olivine and Silica, which are often incorrectly predicted as Pyroxene. Examples of model prediction for each study sample are shown in Figure 4.1.

Mineral	**Accuracy**	**Weighted Accuracy**	**Dice**
Pyroxene	88.7 %		
Plagioclase Feldspar	78.0 %		
Olivine	31.5 %		
Silica	26.9 %		
Matrix	82.0 %		
Overall	81.1 %	61.5%	0.66

Table 4.1: LGBM model test performance.

Use of a Gaussian filter demonstrated the biggest impact on model improvement in pre-processing, improving similarity in image character between PPL, XPL and QEMSCAN inputs. Equalisation techniques had a minimal impact on performance metrics. While this is likely due to use of common equipment in data collection, equalisation is expected to have a significant influence in alternate sampling strategies. Ultimately, the equalisation techniques explored are of limited value in this application, due to the sensitivity in colour representation caused by cropping, and it is recommended that data augmentation should be further considered as an alternative to standardisation.

Feature importance highlights the strong contribution of colour representation in classification performance, both from PPL and XPL images. Edge, ridge and textural features were of lesser influence, but it is not possible to state whether this reflects data characteristics or properties of the underlying mineralogy. Improvement in performance may be gained by better capturing variability of the colour and textural response, through a featuring weighting approach typical of a multilayer perceptron, or through increased flexibility in filter kernels within a fully convolutional neural network.

Concern was highlighted over the potential for over-fitting when using a RF model. This observation was further examined by taking an image from each sample to create an out-of-sample test set and fitting both RF[W] and LGBM models to the remaining dataset. Out-of-sample test accuracy for RF[W] and LGBM were 89.7% and 80.4% respectively, while validation errors for both models were comparable with LGBM test performance (Table 4.2).

Within literature, there are multiple authors proposing use of the RF model in multi-mineral classification. Caution is recommended with use of RF for semantic segmentation. As demonstrated, without applying constraints on tree depth or minimum leaf size, there is a tendency to overfit, which was not immediately obvious without use of an appropriate out-of-sample test data strategy.

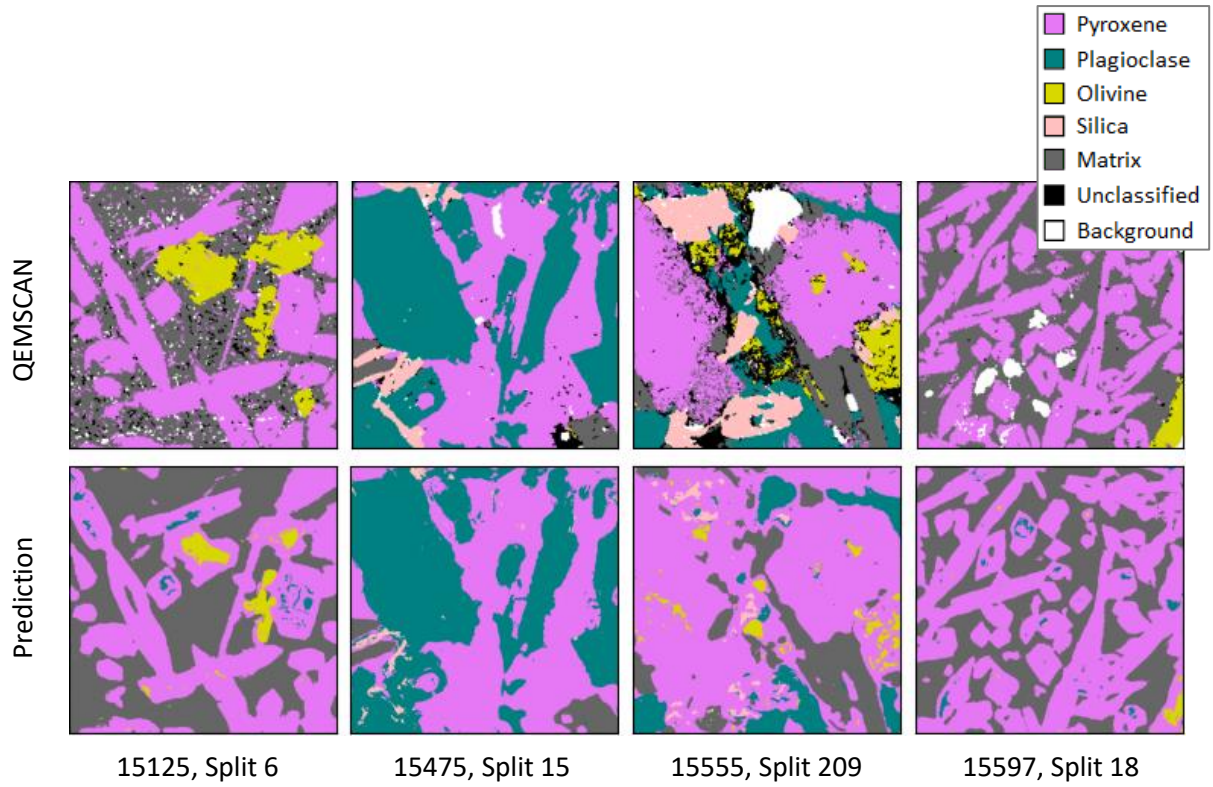


Figure 4.1: Example of LGBM predictions for each study sample.

	Test Performance			**Out-of-sample Test Performance**		
Model	Acc	Acc[W]	Dice	Acc	Acc[W]	Dice
RF[W]	0.897	0.825	0.868	0.799	0.543	0.554
LGBM	0.804	0.602	0.646	0.809	0.555	0.569

Table 4.2: Model test and out-of-sample test performance for weighted Random Forest (RF[W]) and LightGBM (LGBM).

4.2 Mineral Response

To further consider the possibility that model performance is limited by the specific suite of minerals present within this study, a feasibility study is performed to understand the potential for linear discrimination of colour, shape and textural feature responses of study mineralogy.

For the feasibility study, an alternate approach is taken towards feature creation. Taking inputs of PPL, XPL and QEMSCAN interpretation images, a sample of 400 pixels is randomly sampled across all study samples, with equal representation from each class label. The reduction in overhead of feature calculation improves efficiency and allows for rapid scenario testing, such as investigating the impact of classification performance through inclusion of Haralick features (Section 3.3.5).

To quantify the separation of each class within its one-dimensional space, data for each feature must be first be normalised to remove influence of differing features scales, as performed by mean value subtraction and scaling to unit variance. Sample bias towards frequency is compensated by representing each class as a normal distribution, following its central tendency and standard deviation. A Wasserstein distance metric is used to quantify the difference between normalised class distribution. Wasserstein distance is iteratively compared to all other classes within the feature space, with the minimum distance between feature classes proposed as the metric of feasibility.

Feasibility analysis estimated from the study dataset is shown in Figure 4.2. Lower scores indicate an inability to uniquely define a mineral in linear space. The interpretation being that the feature properties of the mineral are not unique, either due to limitations of the physical mineral properties or how they are represented through images and features within the dataset. One potential approach would be to use this information to guide class definition and merge classes. The analysis supports the theory that Plagioclase Feldspar, Olivine and Silica possess overlapping properties requiring use of more complex non-linear relationships if discrimination is possible at all.

Use of this approach creates the opportunity to explore feature design and choice. For example, the PPL CIELAB L attribute generated from is predicted to prove useful in discriminating the matrix classification (Figure 4.3). Feasibility analysis with inclusion of Haralick features indicates there is limited potential for improvement in mineralogy discrimination (Figure 4.4 and 4.5).

As presented, this workflow does not necessarily provide an accurate indication of final model performance. However, the principal deserves further attention, with potential to further improve analysis through inclusion of multi-dimensional analysis.

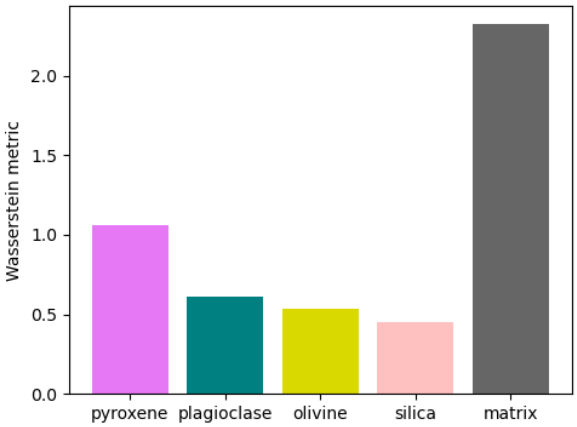


Figure 4.2: Study feasibility based on potential for linear discrimination expressed through maximum Wasserstein distance per feature.

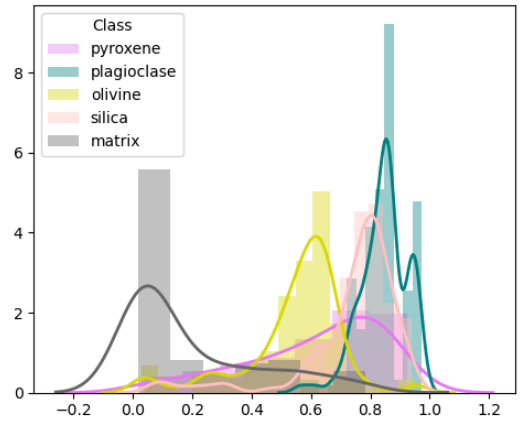


Figure 4.3: Probability density function for PPL CIELAB L attribute.

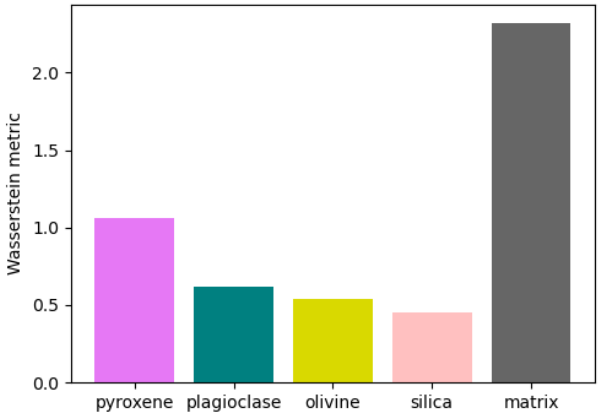


Figure 4.4: Study feasibility including Haralick features based on potential for linear discrimination.

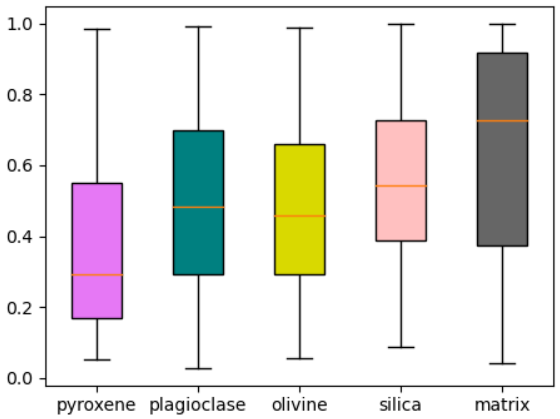


Figure 4.5: Box plot for Haralick homogeneity.

5. Application

An interactive web-application is created to demonstrate potential use of the final classification model. As rock composition rather than mineral classification are of importance in routine rock analysis, the application demonstrates the potential for secondary analysis by using the classification to estimate bulk rock composition and texture properties. Through the graphical user interface sidebar, an end-user may select reference images taken for each sample or upload custom images for analysis, with the inputs and classification shown within the main panel (Figure 5.1). The user may then select to view bulk rock composition or textural property interpretations for the chosen images (Figure 5.2). The textural properties of orientation and roundness are calculated from cluster analysis applied to the classification, performed using Simple Linear Iterative Clustering (SLIC) (Achanta *et al*., 2010), for which parameterisation can be modified within the sidebar. Interpretations are included to demonstrate the scope for common rock analysis, and in a working solution, should be extended to include rock classification, grain-size distribution and elastic behaviour.

The web-application was developed using the python Streamlit app framework. Image processing and feature creation is performed solely through the scikit-image library, to reduce package dependencies and simplify installation and maintenance of the application. Bulk rock composition and texture properties are estimated using scikit-image region properties functionality. Bulk rock composition being the area of each label in the classification, while texture properties reflect cluster orientation and the ratio of length for major and minor cluster axis. The application code is written as a standalone script to maximise performance gains generated from design choices taken in feature and model selections. Application code is provided in a separate digital repository. Details of the repository directory structure are provided in Appendix 9.3.

Figure 5.1: Application: Sample classification using pre-defined images.

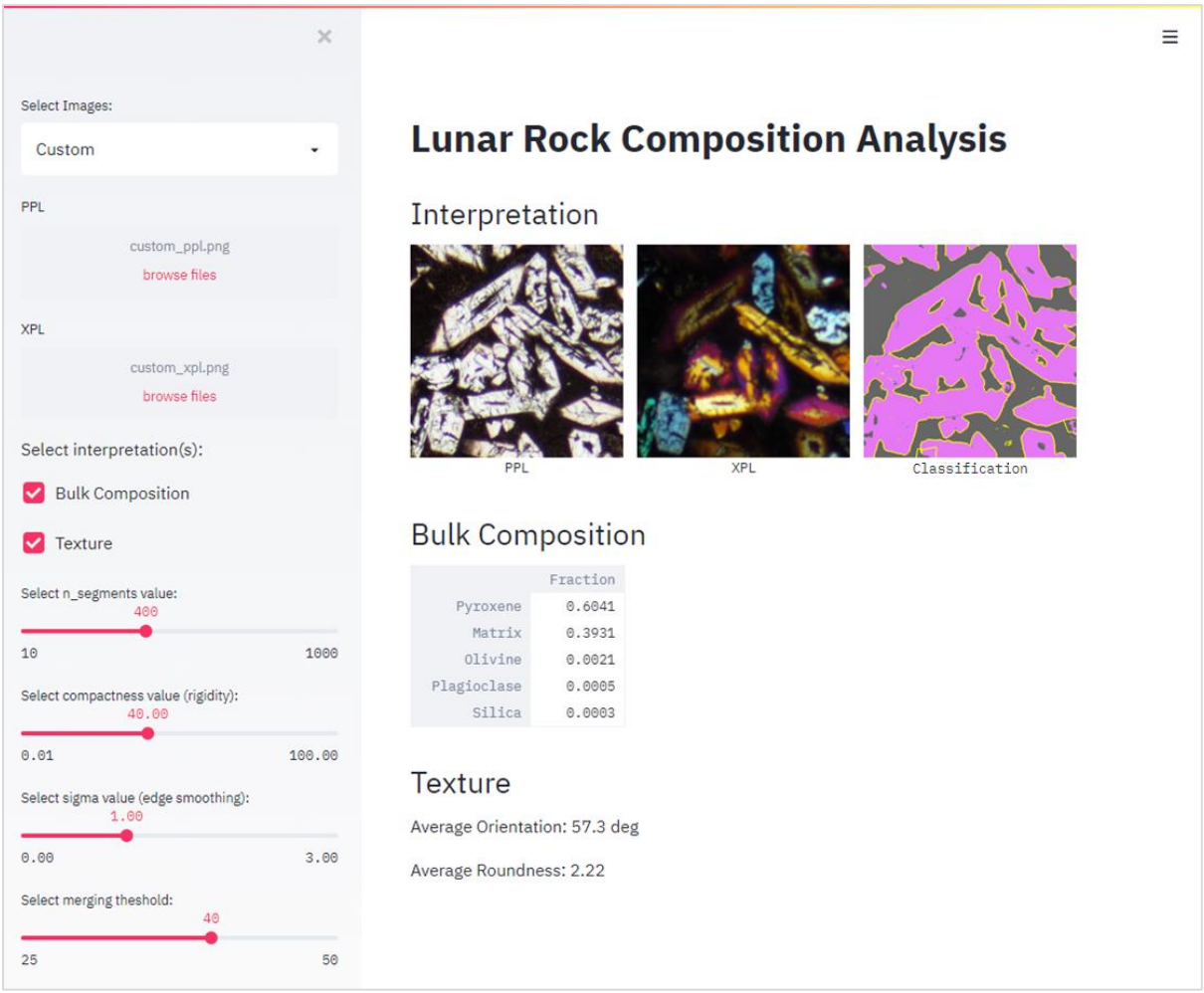


Figure 5.2: Application: Custom image classification with bulk composition and texture analysis. Texture analysis based on clusters fit to classification through SLIC (yellow lines on classification).

6. Conclusions

This study outlines a methodology to perform routine high-level compositional rock analysis of Lunar samples, performed by an end-user though a web-based application. However, the techniques and lessons learned have wider implications in the fields of multi-mineral semantic segmentation, and broader rock classification through machine learning.

A novel workflow has been defined to perform pixel-wise alignment of microscopy and quantitative interpretation from a dataset where collection strategy is not designed for use in machine learning, compensating for variation in rotation, scale, translation and geometric distortion.

A wide variety of features were considered with respect to classifying mineralogy typical of Lunar geology, representing variation in colour, structural and textural response of the minerals captured through microscopy imaging. Key consideration has been given to the understanding of feature properties and their impact on predictive performance, as demonstrated through a feasibility study. Feature analysis indicates colour representation is a primary driver in classification, while texture features provide potential to reduce classification noise.

The selected model uses LightGBM, a gradient boosting tree ensemble implementation, which achieves an overall accuracy of 81.1% and Dice similarity coefficient of 0.66. Overall accuracy is achieved through strong predictive performance within predominant mineralogies, while struggling to classify Olivine and Silica, for which properties are seen to overlap significantly with Pyroxene. Caution has been applied in considering the influence of class imbalance, both in terms of defining a solution and recording performance, and in understanding the influence over-fitting, which can be mitigated using an appropriate validation and hyperparameter tuning strategy.

With focus on producing an end-user web-based application, design choices were taken on classification label selection, feature and model selection to reduce processing overhead, and in not performing image standardisation due to issues introduced by using cropped images. These constraints are not necessarily appropriate in all functions and may provide scope to improve model performance further.

Key challenges faced in the study were found in image processing, due to limited availability of suitable datasets within this field, and through use of a previously untried mineralogical suite. Integrating a data collection phase in similar studies is highly recommended, which at a minimum should ensure consistency in PPL and XPL image alignment, rotation and overlap. While these challenges were ultimately overcome, it is desirable to minimise pre-processing time and to begin an assessment of feasibility, especially where dealing with unproven mineralogy.

7. Future Work

A significant proportion of this study focused on creation of a dataset suitable for use in machine learning, as no open source datasets were available. With new availability of 66.3M pixels of microscopy data aligned to quantitative interpretation data as provided by Mendeley Data, the findings of this study offer an excellent opportunity towards consideration of deep learning approaches in multi-mineral classification with goal of improved mineralogy discrimination and generalisation.

Given the goal for this field is to create an approach to multi- mineral phase identification with proven generally and therefore widespread applicability. To achieve this goal, a comprehensive data collection strategy would be required, or a solution integrated into hardware, such as an optical IoT microscope

Interpretation would be further strengthened by inclusion of domain knowledge, biasing mineral interpretation to obey known relationships of mineral or phase cooccurrence, based around source chemistry, pressure-temperature and chemical stability relationships. Finally, integration between the research areas of multi-mineral classification and texture classification (Singh *et al.*, 2009) provides the potential to create an integrated solution to classify a rock both in terms of its mineralogical and textural characteristics.

8. References

8.1 Publications

Achanta, R., Smith, K., Lucchi, A., Fua, P., Susstrunk, S., 2010, SLIC Superpixels, EPFL Technical Report 149300.

Bell, S., Joy, K., Pernet-Fisher, J., Hartley, M., 2020, Data for: "QEMSCAN as a method of semi-automated crystal size distribution analysis: Insights from Apollo 15 mare basalts", Mendeley Data, v1.

Berrezueta. E., Domínguez-Cuesta, M. J., Rodríguez-Rey, Á, 2019, Semi-automated procedure of digitalization and study of rock thin section porosity applying optical image analysis tools, Computers and Geosciences, 124, 14-26.

Brown, M., Lowe, D. G., 2007, Automatic panoramic image stitching using invariant features, International journal of computer vision, 74, 59-73.

Clark, M., Bovik, A. C., Geisler, W. S., 1997, Texture segmentation using Gabor modulation/demodulation, Pattern Recognition Letters, 6, 261-267.

Dunlop, 2006, Automatic Rock Detection and Classification in Natural Scenes, Master's thesis, Carnegie Mellon University.

Grana, C., Borghesani, D., Cucchiara, R., 2009, Fast Block Based Connected Components Labelling, 16th IEEE International Conference on Image Processing, 4061-4064.

Grove, T. L., Walker, D., 1977, Cooling histories of Apollo 15 quartz-normative basalts, Proceedings 8th Lunar Science Conference, 2, 1501-1520.

Haralick, R. M., Shanmugam, K., Dinstein, I., 1973, Textural Features for Image Classification, IEEE Transactions on Systems, Man and Cybernetics, 3, 610–622.

Izadi, H., Sadri, J., Bayati, M., 2017, An intelligent system for mineral identification in thin sections based on a cascade approach, Computers and Geosciences, 99, 37-49.

Jiang, F., Gu, Q., Hao, H., Li, N., Wang, B., Hu, X., 2018. A method for automatic grain segmentation of multi-angle cross-polarized microscopic images of sandstone, *Computers and Geosciences*, *115*, 143-153.

Longhi, J., Walker, D., Stolper, E. N., Grove, T. L., Hays, J.F., 1972, Petrology of mare/rille basalts 15555 and 15065, The Apollo 15 lunar samples, 131-134.

Lowe, D. G., 1999, Object Recognition from Local Scale-Invariant Features, International Journal of Computer Vision, 60, 91–110.

Maitre, J., Bouchard, K., Paul Bédard, L., 2019, Mineral grains recognition using computer vision and machine learning, Computers and Geosciences, 130, 84-93.

Muja, M., Lowe, D. G., 2012, Fast Matching of Binary Features, Proceedings of the Conference on Computer and Robot Vision, 404–410.

Ojala, T., Pietikäinen, M., Mäenpää, 2002, Multiresolution Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24, 971-987.

Rubo, R. A., Carneiro, C., Michelon, M F., Gioria, R., 2019, Digital petrography: Mineralogy and porosity identification using machine learning algorithms in petrographic thin section images, Journal of Petroleum Science and Engineering, 183.

Schnare, D. W., Day, J. M. D., Norman, M. D., Liu, Y., Taylor, L. A., 2008, A laser-ablation ICP-MS study of Apollo 15 low-titanium olivine-normative and quartz-normative mare basalts, Geochimica et Cosmochimica Acta, 72, 2556-2572.

Singh, N., Singh, T. N., Tiwary, A., Sarkar, K. M., 2009, Textural identification of basaltic rock mass using image processing and neural network, Computers and Geosciences, 27, 301-310.

Thompson, S., Fueten, F., Bockus, D., 2001, Mineral identification using artificial neural networks and the rotating polarizer stage, Computer and Geosciences, 27, 1081-1089.

Weigand, P. W., Hollister, L. S., 1973, Basaltic vitrophyre 15597: An undifferentiated melt sample, Earth and Planetary Science Letters, 19, 61-74.

Wong, S. C., Gatt, A., Stamatescu, V., McDonnell, M. D., 2016, Understanding data augmentation for classification: when to warp?, Computer Vision and Pattern Recognition, arXiv.

8.2 Books

Fairchild, M. D., 2013, Color Appearance Models. John Wiley and Sons. 472 pp.

Nesse., W. D., 2004, Introduction to Optical Mineralogy, Third Edition. Oxford University Press. 348pp.

Raith, M. M., Raase, P., Reinhardt, J., 2012, Guide to Thin Section Microscopy, Second Edition. Open Access Publication. 127pp.

8.3 Internet

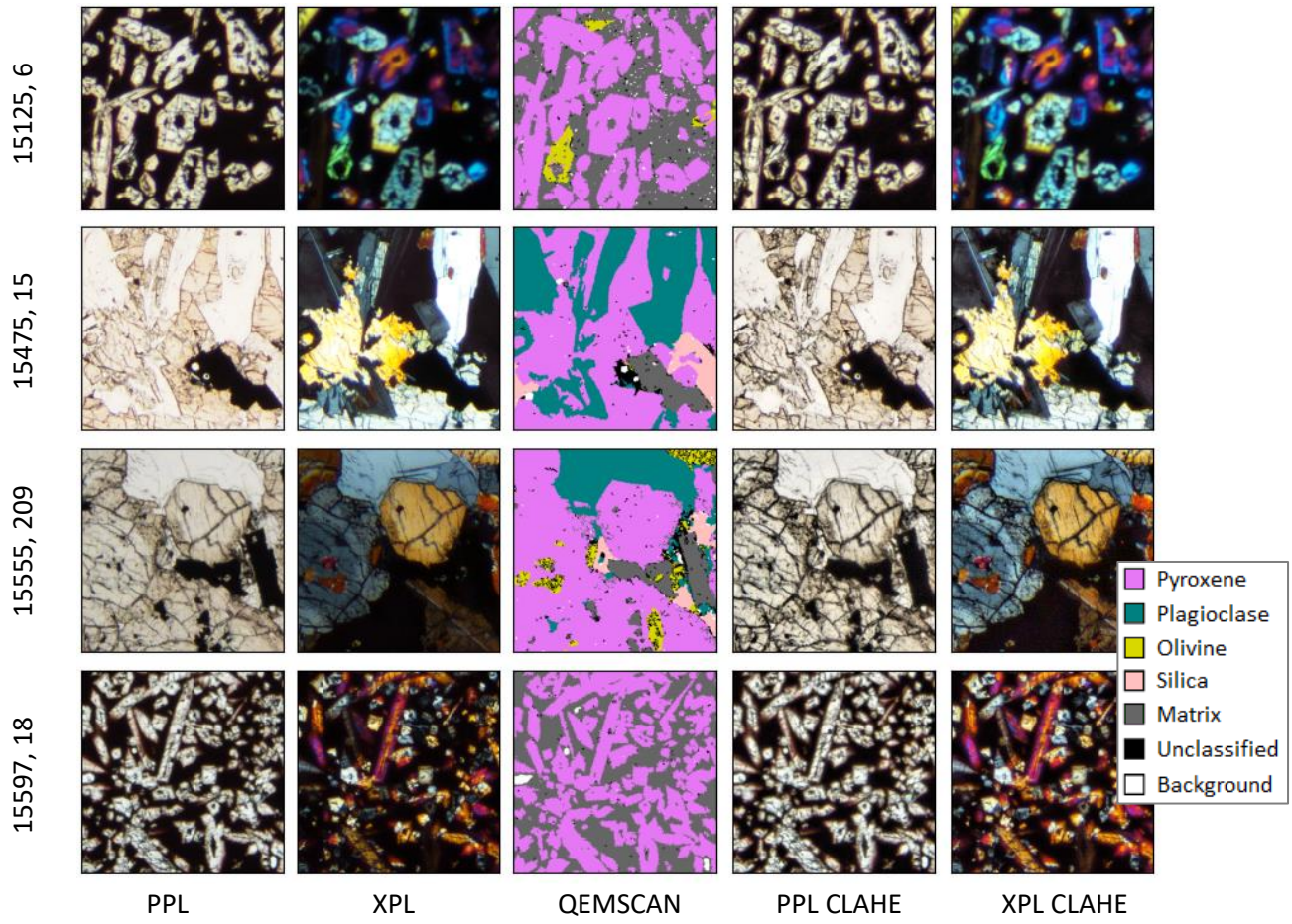
Jónathan Heras (Stack Overflow), https://stackoverflow.com/questions/42459493/sliding-window-in-python-for-glcm-calculation, accessed 14th June 2020.

LPI Lunar Sample Atlas, https://www.lpi.usra.edu/lunar/samples/atlas/#Apollo%2015, accessed 17th April 2020.

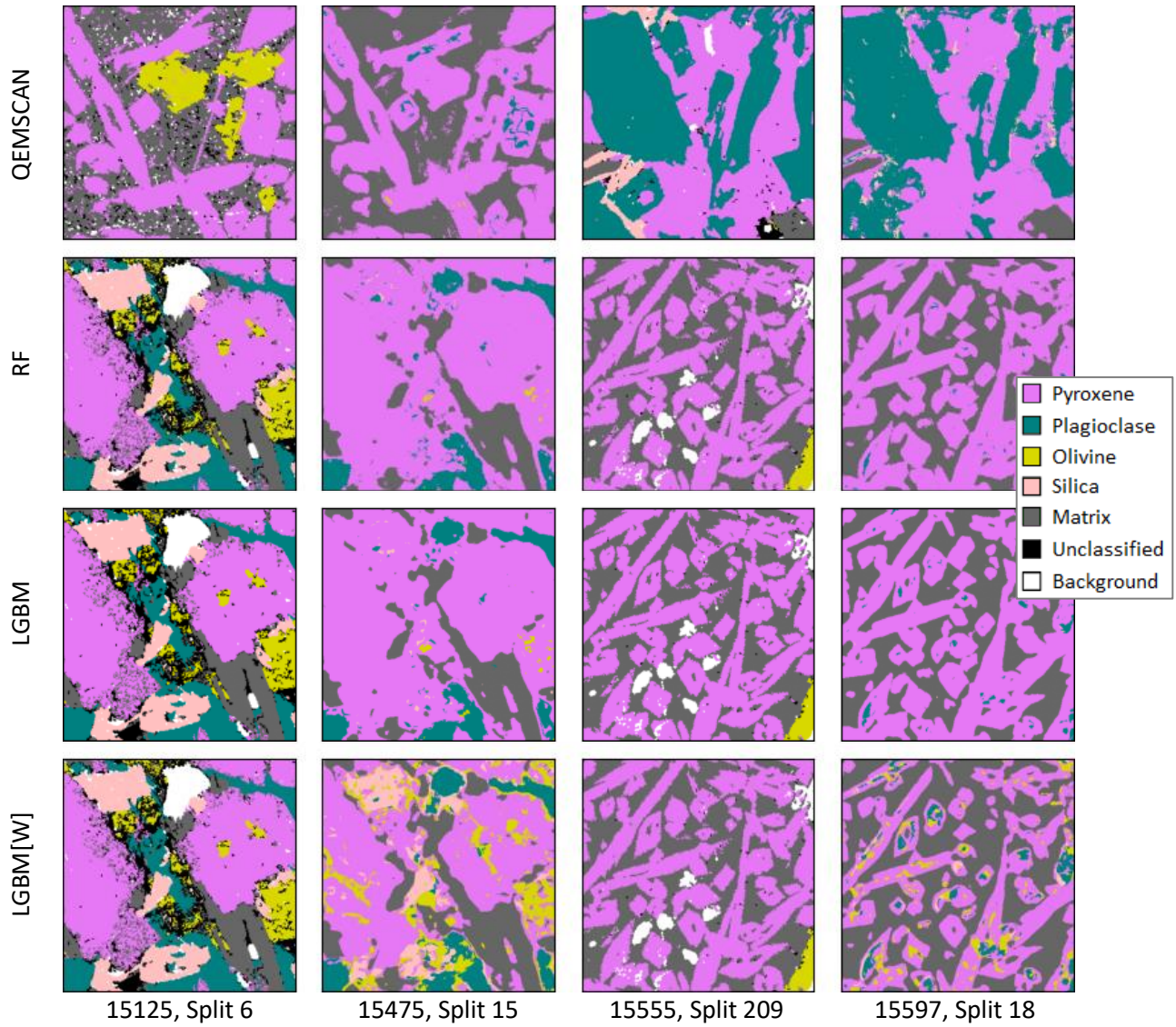
Mendeley Data, https://data.mendeley.com/datasets/rh37cdm9hv/1, accessed 14th April 2020.

NASA Lunar Sample Compendium, https://curator.jsc.nasa.gov/lunar/lsc/index.cfm, accessed 17th April 2020.

9. Appendix



Appendix 9.1: Image adaptive equalisation using Contrast Limited Adaptive Histogram Equalization (CLAHE). Expanded view from **Figure 3.2**.



Appendix 9.2: Model performance on out of sample images. Expanded view from **Figure 3.16**.

```
|  a_preproc.py
|  b_feat_creation.py
|  c_model_selection.py
|  d_feat_feasibility.py
|  haralick.py
|  image_stitch.py
|  README.md
|
|---app
|  |  custom_clf.png
|  |  custom_ppl.png
|  |  custom_xpl.png
|  |  main.py
|  |  requirements.txt
|  |
|  └---images
|        15125_6.png
|        15125_6a.png
|        15125_6c.png
|        15475_15.png
|        15475_15a.png
|        15475_15c.png
|        15555_209.png
|        15555_209a.png
|        15555_209c.png
|        15597_18.png
|        15597_18a.png
|        15597_18c.png
|
|-—data
|  |---15125,6
|  |  |  1.png
|  |  |  1a.png
|  |  |  1b.png
|  |  |  1b1.png
|  |  |  1b2.png
|  |  |  1c.png
|  |  |  1c1.png
|  |  |  1c2.png
|  |  |  2.png
|  |  |  …
|  |  └---inputs
|  |        1.tif
|  |        1x.tif
|  |        …
|  |        contrast_pyx.png
|  |        interp.png
|  |        label.png
|  |        source.png
|  |
|  |---15475,15
|  |  |  …
|  |
|  |---15555,209
|  |  |  …
|  |
|  |---15597,18
|  |  |  …
|  |
|  |---composite
|  |     comp_clf.png
|  |     comp_ppl.png
|  |     comp_xpl.png
|  |
|  └---ref
|        ppl.png
|        xpl.png
|
└---model
     multi_min.clf
```

Appendix 9.3: Digital Code Deliverable Project Structure.