Classifying post-wildfire burn severity in the Kalahari: A methodological approach using drone derived imagery and Random Forest Classification algorithms.

Maddie Gillespie

1. Introduction

Africa is often referred to as the Fire Continent, as African savanna fires account for over 60% of global fire extent (Komarek 1971). The effect of fire on savanna vegetation depends upon the type and intensity of fire, and the season and frequency of burning, i.e. fire regime (Trollope 1999). Fuel load is a key variable affecting fire impact: fires tend to occur more frequently and burn with greater intensity when savannas have a high grass biomass (Trollope 1999). Generally, the main effect of fire on woody vegetation in arid savannas is to cause a topkill of stems and branches, forcing the plants to coppice from the collar region of the stem (Archer et al. 2017). This effect has been seen to positively correlate with fire intensity (Trollope 1999). In the Kalahari, an arid savanna ecoregion within Southern Africa, higher grass fuel loads and a drier climate result in annual late dry season bushfires. Topkill of woody vegetation in the Kalahari is the predominant post-burn response, with coppicing occurring postburn from surviving below ground root structures (Archer et al. 2017 and Holdo 2005). In the Kalahari, discerning tree mortality vs survival post burn is difficult due to surviving below ground biomass that is invisible to the average surveyor. Woody vegetation that has lost significant amounts, if not all above ground biomass, is often fully intact below the ground surface (Holdo 2005). Coppicing of woody vegetation in the Kalahari therefore poses an interesting challenge to researchers and land managers aiming to quantify the post-burn vegetative response.

This research aims to determine the relationship between bush fire severity and post-burn coppicing/regeneration events in the Kalahari Desert in Southern Botswana. To date, there has been a significant lack of understanding or quantification of woody vegetation's response to wildfires of varying intensities in this region. This study will allow a preliminary insight for researchers and land managers into the likelihood of regeneration post-burn, and will therefore allow for estimates of mortality rates and future succession patterns in the landscape. Quantification of post-burn woody vegetation regeneration rates is significant for furthering the understanding of the effect of fire on vegetation composition and community shifts in the Kalahari. This information is necessary to further inform land managers on the effects of differing fire regimes on vegetation response in the context of management policies such as prescribed burning and shrub encroachment control. This report will outline the initial methodological phases of this proposed research, specifically detailing the usage of machine learning classification algorithms to classify burn severity post burn from drone derived imagery. The proposed model for classifying burn severity will then be implemented in later analyses to discern the relationship between the severity rankings and post burn regeneration.

Burn severity is best defined as a measure of fire effects, which can be described by the degree of mortality in aboveground vegetation, or by the degree to which the fire alters soil properties and below ground processes (Smith et al. 2005 and Keeley 2009). Previous studies have classified burn severity using spectral parameters derived from multi- and hyperspectral satellite and drone imagery. A common method of classifying burn severity with spectral data is to compute the difference between the pre-fire and post-fire Normalized Burn Ratio (NBR), which is based on the near-infrared and shortwave infrared bands from 30 meter resolution Landsat imagery (Fraser et al. 2017). A previous study by Smith (2005) details the usage of NBR indices to derive burn severity classifications in Northern Botswana with results showing difficulty to accurately classify burn severity, particularly white ash deposits, given the rough spatial resolution of the Landsat imagery comparative to the fine scale ash deposits. In burn severity classifications outside of Southern Africa, there is an underlying assumption that surface albedo decreases following a fire. In savanna fires, two ash endmembers primarily occur, white 'mineral' ash where fuel has undergone complete combustion, and darker 'black' ash or char where an unburned fuel component remains (Hudak et al. 2013). Most burn severity maps focus on darker color as an indicator of higher intensity, but in savannas, white ash is the primary indicator of higher intensity (Smith et al. 2005). The NBR based methodology focuses on black charred surface as the highest indicator of burn severity which does not accurately represent savanna ecosystems response post burn.

Driven by a need to acquire finer spatial resolution scales within burn severity classification, unmanned aerial vehicles (UAVs) offer a means of conducting inexpensive, on-demand remote sensing that can detect the fine scale white ash deposits. UAVs are a promising tool for evaluating spatial variations in burn severity at very high spatial resolution, and can be acquired on demand at a low economic cost. UAV imagery provides more information in terms of spatial variability in heterogeneous burned areas in comparison with high-resolution satellite imagery (Beltrán-Marcos et al. 2021). Within the Kalahari and Southern Africa as a whole, there has been no attempt at using UAV imagery to classify burn severity, with previous research primarily focusing on using satellite imagery for classification. This study will use low cost RGB drone imagery as the primary data source within burn severity classification, a first for the ecoregion.

This study will implement the Machine Learning (ML) Random Forest Classification Algorithm to employ in the burn severity classification. ML techniques have been documented to perform better than simple classifiers in dealing with complex interactions between scene complexity and scale and have improved discrimination of classes in heterogeneous landscapes (which is common in remote sensing) with low inter-class separability and high intra-class variability (Collins et al. 2016). Random Forest Classification in particular has the following advantages in classification mapping: handles categorical predictors naturally, computationally simple to fit, has no formal distributional assumptions, and performs automatic variable selection (Rodriguez-Galiano et al. 2011). The combination of drone derived imagery and Random Forest Classification will be used to create a robust model that can be implemented by researchers and land managers alike to classify burn severity in the Kalahari.

1.1 Study Site

This study will focus on the Kalahari Desert in Southern Botswana, an arid savanna ecosystem. Rainfall is highly seasonal, with an annual mean of 650 mm, which falls almost exclusively between October and April (Meyer et al. 2019). The dry season is divided into cool (May-August) and hot (September-November) seasons (Meyer et al. 2019). Fire is a common type of disturbance, and tends to occur late in the dry season when grasses have low moisture content and ignite readily (Archer et al. 2017). The study was conducted at the Modisa Wildlife Reserve in southern Botswana where a large bush fire burnt a majority of the reserve in the late dry season of 2021. The Modisa Reserve serves as an ideal study site given its lack of land degradation and shrub encroachment due to properly managed game stocks and lack of overgrazing that allows it to represent a healthy savanna ecosystem. The Kgalagadi Transfrontier Fire (Figure 1) was a lightning strike ignited fire that began in August 2021 and ended in late September of 2021, burning over 4,000,000 hectares of land (Kaduyu et al. 2023). A current study in review (Kaduyu et al. 2023) used the LandSat Fire Mapping Tool to estimate total burn severity for the Kgalagadi fire extent, overall claiming low severity estimates for the fire. This study primarily focused on large scale, spectral based severity classifications rather than fine scale drone based methodologies that have been seen in previous studies to better fit the fine scale resolution needs of post-burn mapping in the Kalahari (Smith et al. 2005).

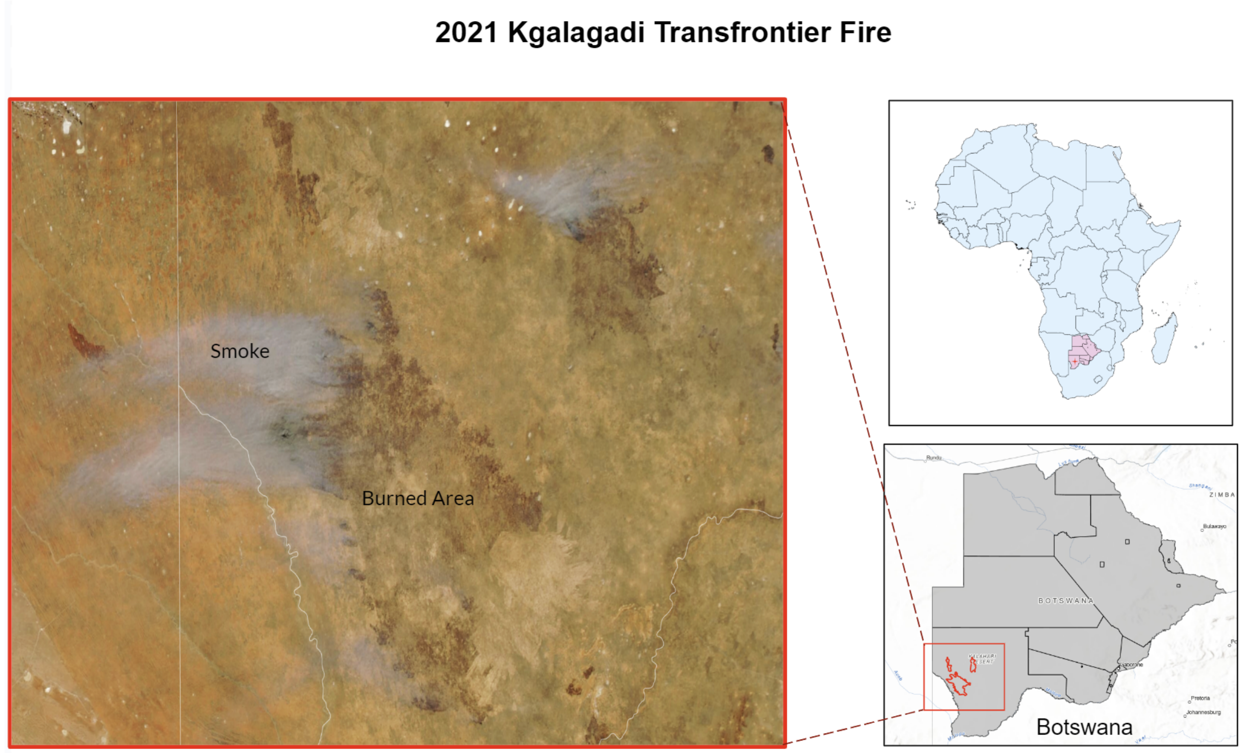

Figure 1: Kgalagadi Transfrontier Fire extent and location in Botswana. Image Source: https://earthobservatory.nasa.gov/images/148829/a-fast-moving-fire-in-kgalagadi

2. Data

2.1 RGB Drone Imagery

A collection of RGB drone derived imagery acquired from the post burn site at the Modisa Reserve is the primary data source used within this research. Post burn imagery was captured over the same 1x1 square kilometer area over the reserve during the following time periods:

Table 1: Date of Image Acquisition & Time Since Fire

Sept 26th, 2021 - 12 hours post-burn
December 29th, 2021 - 6 months post-burn
July 21st, 2022 - 1 year post-burn
August 9th, 2023 - 2 years post-burn
November 23, 2023 - 2.5 years post-burn

The imagery was captured using a DJI Phantom 3 UAV and processed into 5x5 cm resolution orthomosaics in the drone imagery processing software program DroneDeploy. The drone imagery from the 12 hours post-burn imagery was flown at an early hour in the morning, resulting in an increase of shadows in the image. Shadow was therefore classified as its own class in an attempt to separate the shadows and minimize their effects. Images were georeferenced to the study site within DroneDeploy and DEM and DSM images were created using point cloud data within the software. For this preliminary analysis, the 1x1 sq. km drone image was reduced to a smaller 3,500 sq. meter image to allow for timely data processing and analysis given the lack of computing power within the laptop used for the analysis. The 12 hours post burn image is being used as the primary training and test image. The imagery from years post burn will be used in phase two of this research to link burn severity to post burn coppicing likelihoods. The test drone image was equalized prior to analysis, a processing technique that is used to increase the contrast in images. It accomplishes this by effectively spreading out the most frequent intensity values, i.e. stretching out the intensity range of the image (Nixon and Aguado 2020). However, when the model was initially run, the equalized image yielded a lower overall classification accuracy compared to the raw RGB image. Due to this, the raw RGB image model results were used within the study and final classifications.

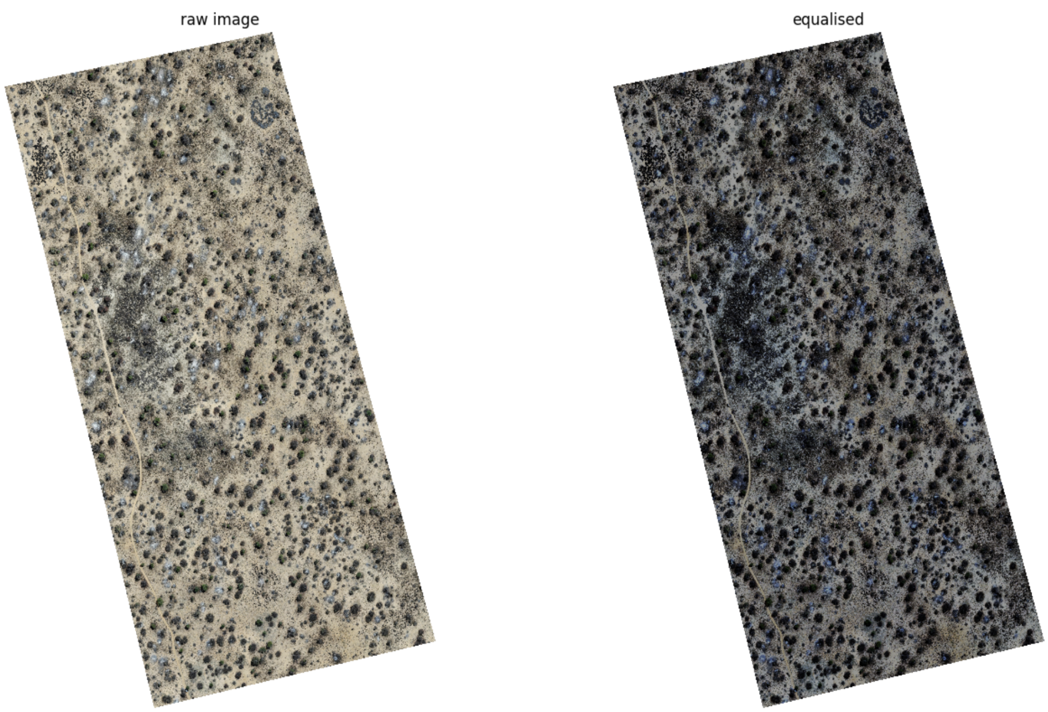

Figure 2: Unprocessed RGB image (left) and the equalized image (right).

2.2 Manual Classification

Given the fine resolution of UAV imagery, biophysical parameters rather than spectral parameters were used as indicators for burn severity. The effects of the fire in the drone imagery are easily discernible by the human eye in the fine scale imagery, and therefore it was determined that a manual classification of land cover classes that corresponded to specific burn severity rankings was the best method for ensuring the highest level of accuracy in the training dataset. Many previous studies classifying burn severity have used spectral indices as indicators of burn severity, but given the high resolution of drone imagery, this study aims to test the accuracy of using biophysical indicators that allow for easy interpretation and manual classification (Hillman et al. 2021). As seen in Figure 3 below, physical indicators such as green vegetation and white ash deposits are clearly visualized. The training data for the model was therefore decided to be based upon manually classified land cover classes that correspond to burn severity rankings as detailed in Table 2.

Figure 3: A zoomed in excerpt of the full RGB drone image where differences between fire effects is clearly visible, as seen with the green vegetation vs white ash.

Table 2: Land Cover Classification Schema & Burn Severity Rankings

Classification Schema	**Severity Ranking**
Green Vegetation	1 - Low Severity
Bare Soil	1 - Low Severity
Burnt Vegetation (Still Standing & Uncharred Grass)	2 - Med Severity
Gray Ash	3 - High Severity
Charred Soil/Vegetation/Woody Vegetation	3 - High Severity
White Ash	4 - Total Combustion

Manual land cover classifications were performed in ArcPro in which a subset dataset was created that included two segmented images from the test RGB image and then the corresponding manually classified images for the areas (Figure 5). Feature datasets and classes were defined in ArcPro as seen in Figure 4 that correlate with the listed classification schema in Table 2 above. These two manually classified images and their corresponding RGB images were used to train and evaluate the Random Forest Classification model.

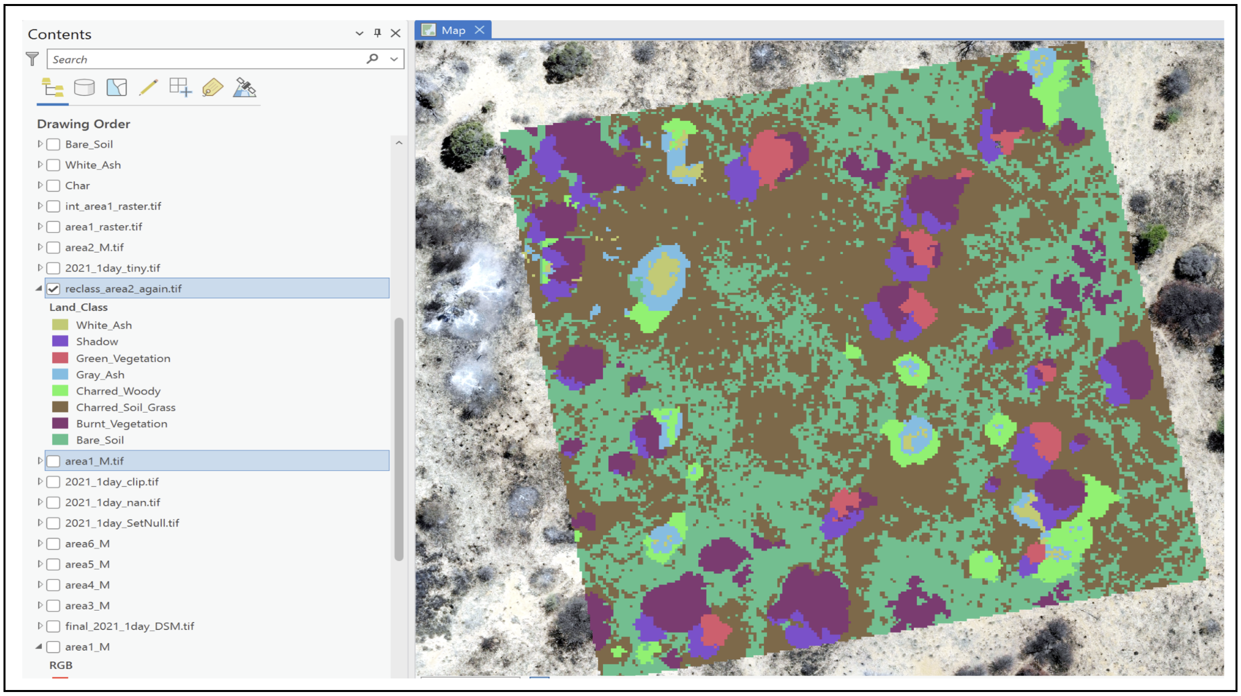

Figure 4: Manual classification set up in ArcPro and finalized manually digitized image.

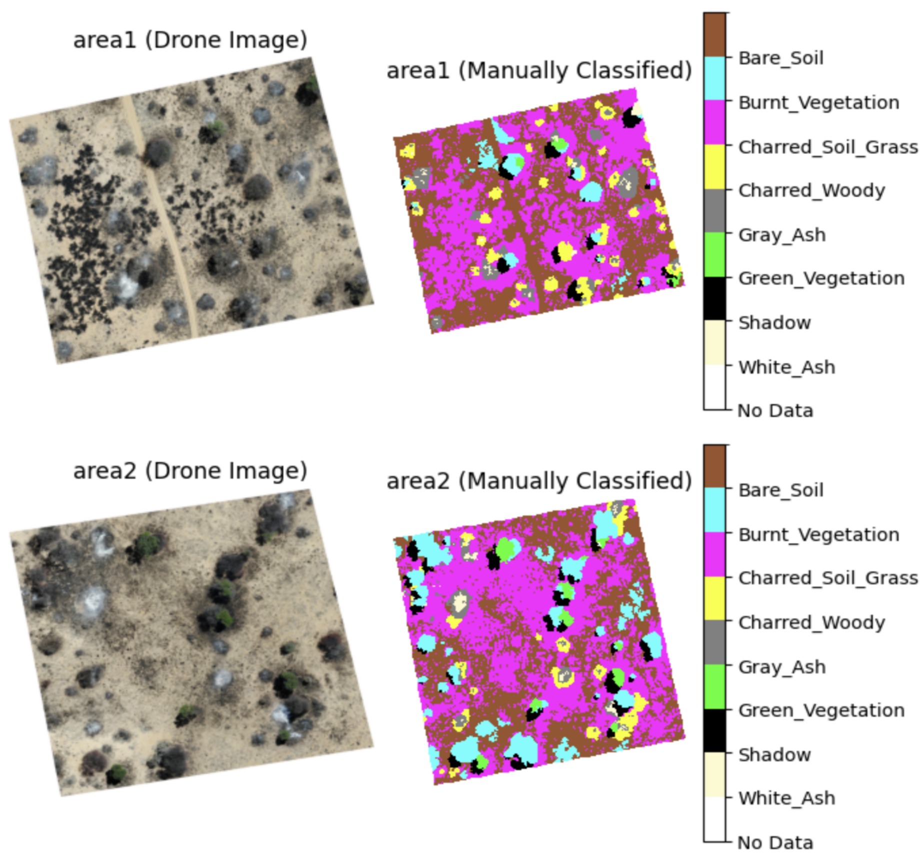

Figure 5: Training and evaluation images: Raw RGB drone images (left) and the manually classified land classification data (right).

3. Modeling Methodology

Random Forest Classification has often been used in land cover classifications due to its relatively stable and robust classification accuracy and effectiveness in handling large and high-dimensional datasets. During the pre-analysis planning phase, a K-Means clustering algorithm was considered for the classification, but it was decided that a supervised classification with clearly defined manually classified data was the best option to ensure high accuracy. An unsupervised classification for this research is a viable path, but given the high resolution of the drone imagery that allowed for easy separation of the desired classes, a supervised classification was preferable.

For model training, the typical ML approach was applied in which 60% of the data points were selected while the remaining 40% of data points were split to create the "testing" data, used to unbiasedly evaluate the model's fit on the training dataset. Within this model, two subsets of the overall RGB drone image and their corresponding manually classified images (Figure 5) were used to train and evaluate the model, with the RGB data set as the 'x' dataset and the manual classifications as the 'y' dataset. The hyperparameters seen in the code block in Figure 6 were used when implementing the Random Forest Classifier. A bootstrap parameter was used to increase variability and diversity during the construction of individual trees in the model, helping prevent overfitting and improving its robustness and generalization to different land cover classifications (Rodriguez-Galiano et al. 2011). Even with bootstrap parameters, random forests can still be prone to overfitting, especially if the number of trees in the forest is very large. While they are less prone to overfitting than individual decision trees, it's essential to tune hyperparameters such as the maximum depth of the trees and the number of features considered at each split (Rodriguez-Galiano et al. 2011).

Evaluation metrics including overall classification accuracy, precision, recall, F1 scores, and support (number of pixels) were calculated when evaluating the model to provide further statistical information on the accuracy of the model. Precision is defined as the number of true positives over the number of true positives plus the number of false positives (sklearn.metrics). Precision can best be defined as the ability of the classifier not to label as positive a sample that is negative. Recall is defined as the number of true positives over the number of true positives plus the number of false negatives (sklearn.metrics). Recall can best be defined as the ability of the classifier to find all the positive samples. These quantities are also related to the F1 score, which is defined as the harmonic mean of precision and recall (sklearn.metrics). For each class, it gives an overall indication of performance, where precision and recall are weighted equally.

```
RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
                       max_depth=None, max_features='auto', max_leaf_nodes=None,
                       min_impurity_decrease=0.0, min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, n_estimators=10, n_jobs=-1,
                       oob_score=False, random_state=None, verbose=0,
                       warm_start=False)
```

Figure 6: Hyperparameters used within Random Forest classification model.

4. Results

4.1 Evaluation of the Model

The model had a relatively acceptable overall classification accuracy of 0.73, but still leaves room for future hypertuning and improvement in classification. Table 3 below shows the calculated model statistics for the overall model and each evaluated land classification.

Table 3: Model classification statistics

Land Cover Classification	Precision	Recall	F1-Score	Support
White Ash	0.49	0.38	0.42	20514
Shadow	0.67	0.43	0.52	127851
Green Vegetation	0.74	0.49	0.59	76469
Gray Ash	0.58	0.49	0.53	124476
Charred Woody	0.39	0.18	0.25	224694
Charred Soil & Grass	0.62	0.77	0.69	1595044
Burnt Vegetation	0.47	0.13	0.21	748384
Bare Soil	0.68	0.84	0.75	1359253
Micro average	0.63	0.62	0.63	4286685
Macro average	0.58	0.46	0.50	4286685
Weighted average	0.60	0.62	0.59	4286685

Total classification accuracy: 0.725526

Bare soil and charred soil/grass had the best overall precision and recall values indicating higher levels of accuracy than the other land cover classifications. This can be attributed to the high levels of pixels within the two individual classes seen in the support variable that offered a larger training sample size. Land classes with lower total pixel counts (support) can be seen to have lower precision and recall scores generally. Green vegetation displays an interesting discrepancy between precision and recall, having the highest overall precision (0.74) and a comparatively lower recall (0.43). This indicates that the model rarely classified green vegetation incorrectly, but struggled to detect all of the green vegetation pixels within the image. A high sample size of green vegetation within the training dataset may be necessary to increase the overall recall of this classification.

Figure 7 below shows the calculated confusion matrix for the model, further highlighting the variability in model classification for the different land cover classes. As seen in the confusion matrix, shadow was misclassified often for charred grass/soil, with 0.43 being reported as charred grass/soil. This was an expected issue and further hypertuning of the model will need to account for shadows through a masking method. Charred woody and charred grass/soil were often confused, with a reported score of 0.56 charred woody being classified as charred grass/soil. Future model adjustments should be made to combine the classes in an aim to increase model accuracy. Classes such as white ash, green vegetation, and gray ash performed overall poorly in the model's classification; this can primarily be attributed to the smaller sample size of pixels within these classes.

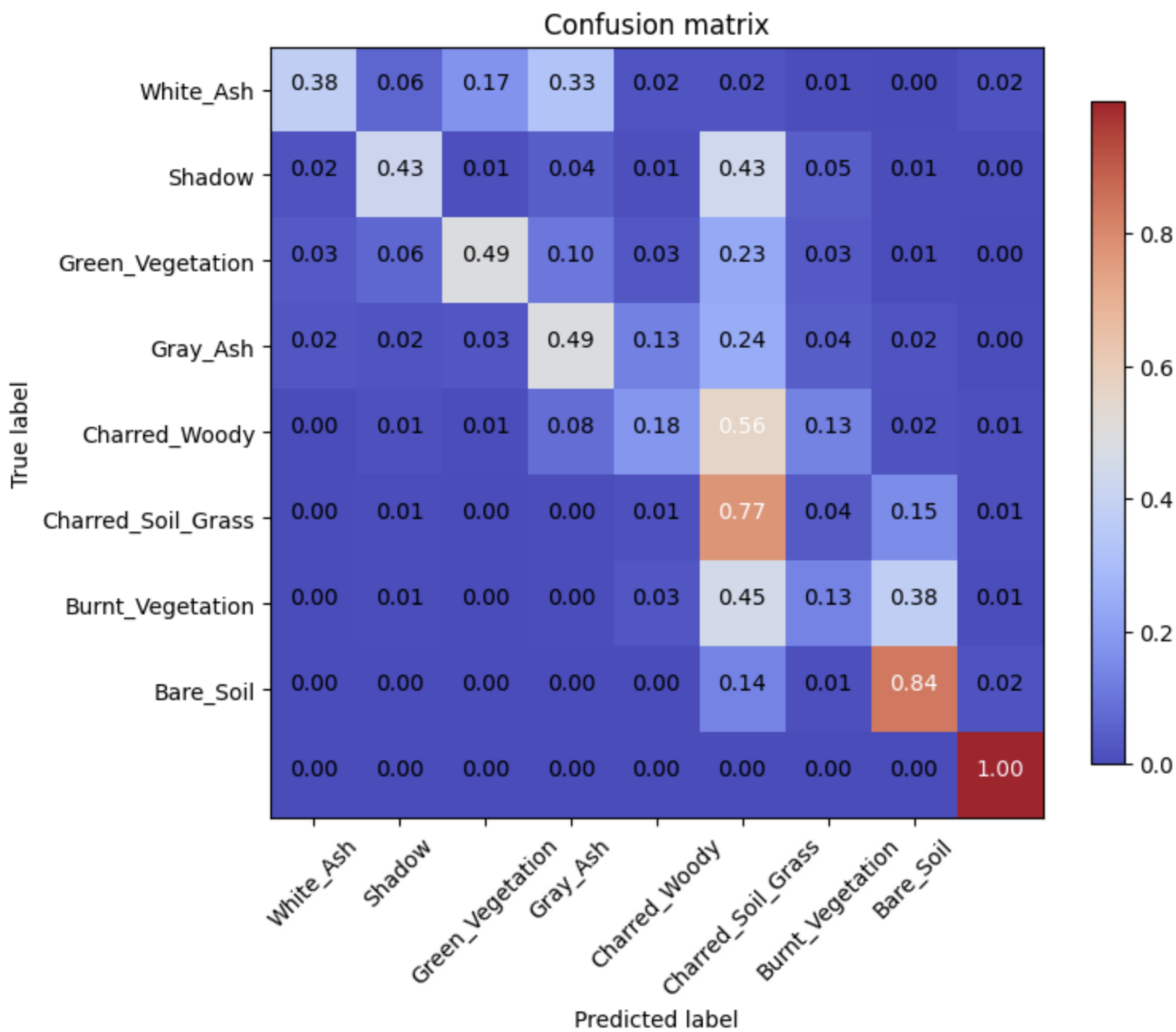

Figure 7: Confusion matrix for the classification. The values along the diagonal in a normalized confusion matrix are the same as the recall scores in the classification report above.

4.2 Predictions

The model was fed the test image and both land classification and burn severity classification maps were produced as seen in Figure 8. Based on the classification of burn severity, it can be seen that the overall burn severity, while patchy and variable across the study site, is relatively high. More model hypertuning and adjustments need to be made for accurate burn severity conclusions, but the results largely match the physical indicators seen in the initial RGB drone image.

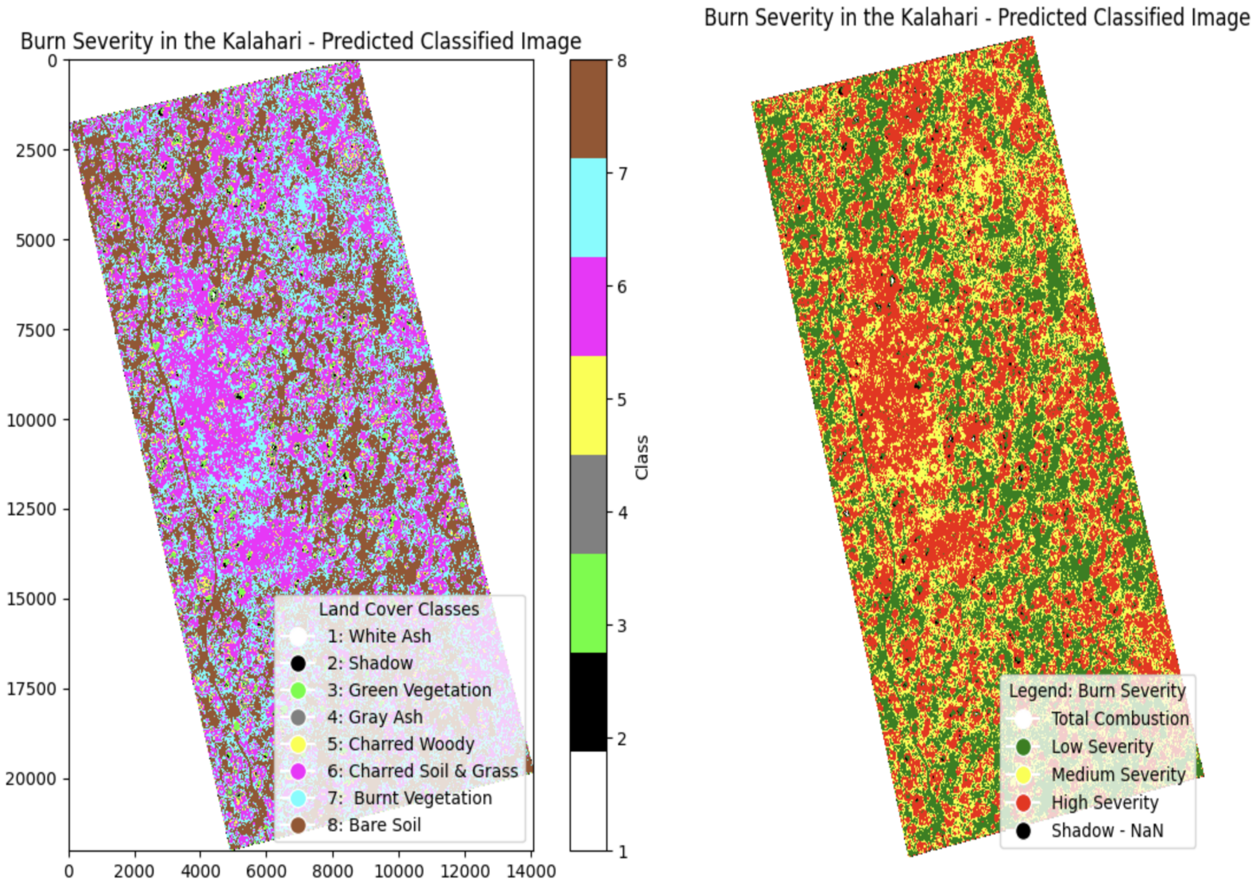

Figure 8: Model predicted land cover classification over the entire drone image (left) and the corresponding burn severity classification (right).

5. Discussion & Future Steps

Overall for a preliminary analysis, the model shows acceptable classification accuracy and with some adjustments, promising future results. For the creation of a fully robust model, certain parameters need to be fine tuned including an increase in both the amount and variation of manually classified land cover training data. An increase in the sample size for land cover classes such as green vegetation, white ash, and gray ash would significantly improve the model. Given that these classifications are less common responses to fire, one the lowest end of the spectrum and the other the highest in terms of burn severity, it is understandable that these classes have the smallest sample size in a larger area. Smaller subsets of manually classified data focusing on these undersampled regions should be added to the training dataset. Imbalanced datasets is one of the common limitations seen in random forest classification where certain land cover classes are underrepresented. It's important to balance the dataset or use class weights to avoid biased predictions. Further adjustments include masking shadows in image preprocessing and processing the model on a computer with higher memory and computing power to ensure the entire drone imagery dataset can be used. During image preprocessing, canopy height models were derived from DSM and DTM. Implementing these canopy height models with certain land cover classes such as the standing burnt vegetation would help increase classification accuracy.

When preprocessing initial imagery, an equalized image was created to normalize the RGB image as it had been suggested in the literature (Nixon and Aguado 2020) to have a positive effect on overall classification accuracy. However, this was not observed in the initial model run through - the equalized image yielded a classification accuracy of 0.68 and the RGB image a classification accuracy of 0.72, therefore the RGB image was used for the final classification and within the final model report. This possibly could be attributed to the intensity of the contrast that made classes such as green vegetation darker and therefore more difficult to distinguish. A path to explore in future model iterations is to normalize the image using HSV color space (hue, saturation, value) rather than RGB because HSV corresponds better to how people experience color than the RGB color space does. Current model fine tuning is exploring this avenue of normalization.

Overall this model shows strong potential for future burn severity classification. With additional hypertuning and adjustments to the training dataset the model will be able to yield higher classification accuracy results. This preliminary analysis and model has the potential to be implemented in post burn assessments throughout Southern Africa, allowing an easy methodology for land managers and researchers to derive vegetation and soil burn severity estimates.

6. References

1. Archer, S. R., E. M. Andersen, K. I. Predick, S. Schwinning, R. J. Steidl, and S. R. Woods. 2017. Woody Plant Encroachment: Causes and Consequences. Pages 25–84 in D. D. Briske, editor. Rangeland Systems. Springer International Publishing, Cham.
2. Beltrán-Marcos, D., S. Suárez-Seoane, J. M. Fernández-Guisuraga, V. Fernández-García, R. Pinto, P. García-Llamas, and L. Calvo. 2021. Mapping Soil Burn Severity at Very High Spatial Resolution from Unmanned Aerial Vehicles. Forests 12:179.
3. Collins, L., P. Griffioen, G. Newell, and A. Mellor. 2018. The utility of Random Forests for wildfire severity mapping. Remote Sensing of Environment 216:374–384.
4. Fraser, R., J. Van Der Sluijs, and R. Hall. 2017. Calibrating Satellite-Based Indices of Burn Severity from UAV-Derived Metrics of a Burned Boreal Forest in NWT, Canada. Remote Sensing 9:279.
5. Holdo, R. 2005. Stem mortality following fire in Kalahari sand vegetation: effects of frost, prior damage, and tree neighborhoods.
6. Hillman, S., B. Hally, L. Wallace, D. Turner, A. Lucieer, K. Reinke, and S. Jones. 2021. High-Resolution Estimates of Fire Severity—An Evaluation of UAS Image and LiDAR Mapping Approaches on a Sedgeland Forest Boundary in Tasmania, Australia. Fire 4:14.
7. Hudak, A. T., R. D. Ottmar, R. E. Vihnanek, N. W. Brewer, A. M. S. Smith, and P. Morgan. 2013. The relationship of post-fire white ash cover to surface fuel consumption. International Journal of Wildland Fire 22:780.
8. Kaduyu, I., R. Tsheko, J. H. Chepete, and E. Kgosiesele. 2023. Burned Area Estimation and Severity Classification Using the Fire Mapping Tool (Fmt) in Arid Savannas of Botswana, a Case Study – Kgalagadi District. preprint, SSRN.
9. Keeley, J. E. 2009. Fire intensity, fire severity and burn severity: a brief review and suggested usage. International Journal of Wildland Fire 18:116.
10. Komarek, Edwin V. "The natural history of lightning." 1971. Proceedings of the Tall Timbers Fire Ecology Conference. Vol. 3.

11. Mark S. Nixon, Alberto S. Aguado, 2020. 3 - Image processing, Editor(s): Mark S. Nixon, Alberto S. Aguado, Feature Extraction and Image Processing for Computer Vision (Fourth Edition), Academic Press, Pages 83-139.
12. Meyer, Holloway, Christiansen, Miller, D'Odorico, and Okin. 2019. An Assessment of Multiple Drivers Determining Woody Species Composition and Structure: A Case Study from the Kalahari, Botswana. Land 8:122.
13. Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 67:93–104.
14. Smith, A. M. S., M. J. Wooster, N. A. Drake, F. M. Dipotso, M. J. Falkowski, and A. T. Hudak. 2005. Testing the potential of multi-spectral remote sensing for retrospectively estimating fire severity in African Savannahs. Remote Sensing of Environment 97:92–115.
15. Scikit Learning - Precision, Recall, Fscore, Support. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html
16. Trollope, W. S. W. 1999. EFFECTS AND USE OF FIRE IN SOUTHERN AFRICAN SAVANNAS. Department of Livestock and Pasture Science, Faculty of Agriculture University Fort Hare, Alice, South Africa.