Seasonal Change of Water Area in Lake Chilwa, Malawi

1. Introduction

Lake Chilwa, the second-largest lake in Malawi, experiences significant seasonal changes in water area [1]. As part of global environmental monitoring, UNEP-WCMC uses supervised classification to distinguish water from other land cover types, including wetlands, vegetation, and cultivated areas. This study offers key insights for local environmental protection and management.

2. Data and Analysis

Four Level-2 Landsat-8 scenes were provided, with two for each date to ensure full coverage of Lake Chilwa[2]. The USGS Collection 2 Level-2 data was already atmospherically corrected[3], so no further processing was needed. The key steps included (**Figure 1**):

1) Image Enhancement and Band Combinations
   i. Mosaicking: Combined two scenes per date for full-lake area coverage.
   ii. Band Enhancement: Applied a 2% stretch to band values.
   iii. Band Selection: Used True Colour (Bands 4-3-2), False Colour 1 (Bands 5-4-3), False Colour 2 (Bands 7-5-4), NDWI, and NDVI to help sampling.
   iv. Clip the layer: Focused on the lake and surrounding area to reduce computation.
2) Secondary data sources (Google Earth, CORINE Land Cover) and related literature are used to collect samples for reference. Stratified random sampling ensured balanced representation across classes [4]:
   i. Waterbodies: High NDWI, low NDVI.
   ii. Wetlands: Relatively high NDVI and NDWI.
   iii. Vegetation: Herbaceous and forested areas appear red in False Colour 1.
   iv. Planted / Cultivated area: Light green or brown areas in False Colour 1.
3) Supervised Classification:
   i. Random Forest (RF): 50 trees, max depth of 30, 1000 samples per class.
   ii. Support Vector Machine (SVM): Tested with 200, 500, and 1000 samples per class.
4) Accuracy Assessment: Evaluated using Precision (User's Accuracy), Recall (Producer's Accuracy), and Overall Accuracy (Correctly classified pixels / Total pixels).

3. Results

The SVM classification results (**Figure 2**) show significant changes in water extent:

1) May 2018: Water coverage is larger, consistent with true color image. Wetlands are distributed along the lake, while surrounding areas are cultivated land and vegetation. September 2018: Water extent significantly reduced, with an increase in wetland areas.
2) Accuracy Assessment (**Table 1**):
   a) SVM performed better than RF, with an overall accuracy of 99.76% on September 19, compared to RF's >95% on both dates.
   b) Water class accuracy: User's accuracy >98% and Producer's accuracy ≈100%, indicating misclassification and omission are minimal.
   c) Note: Increasing the maximum number of samples per class of SVM from 500 to 1000 reduced the accuracy; Therefore, 500 pounds was used.

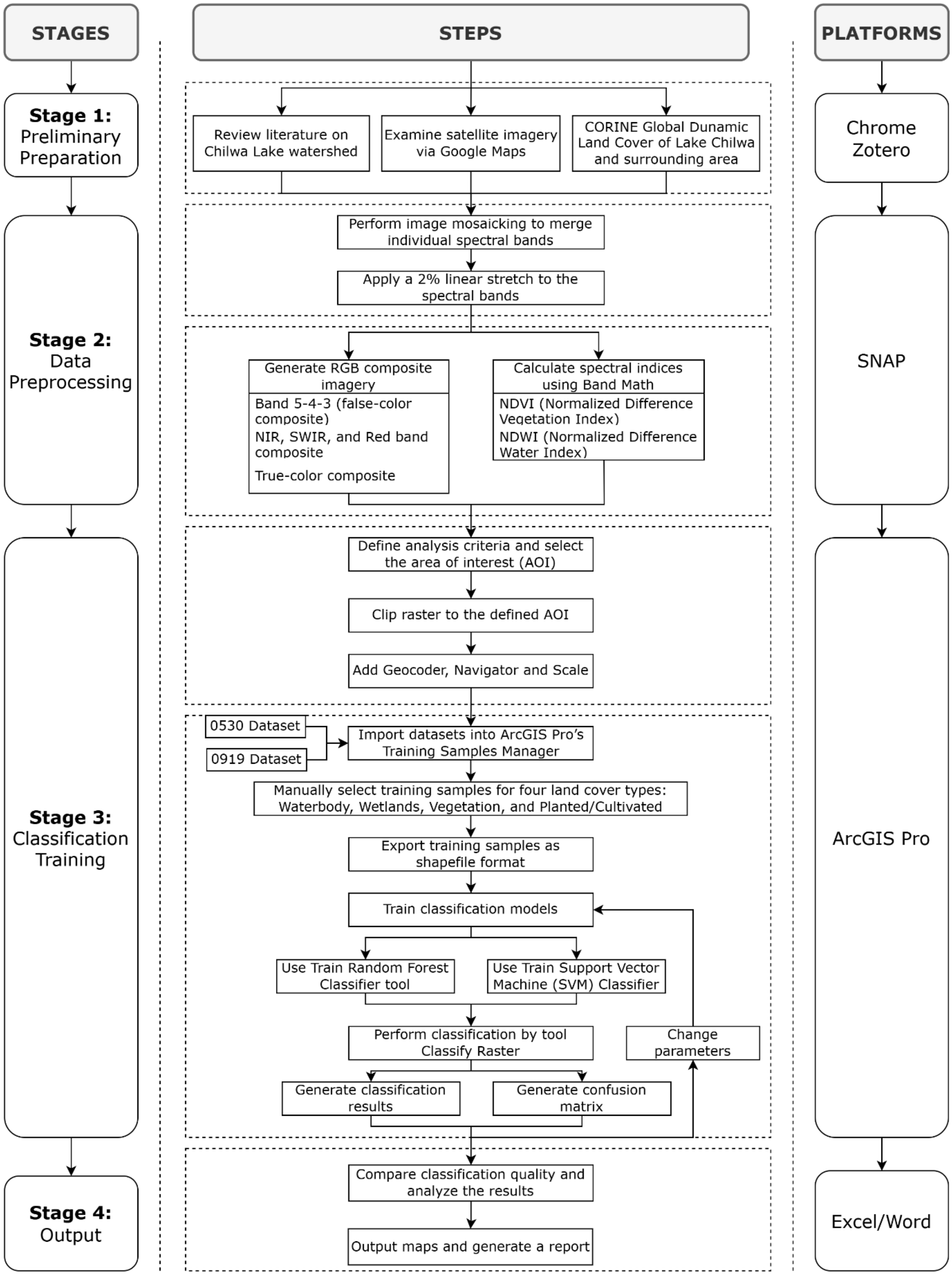
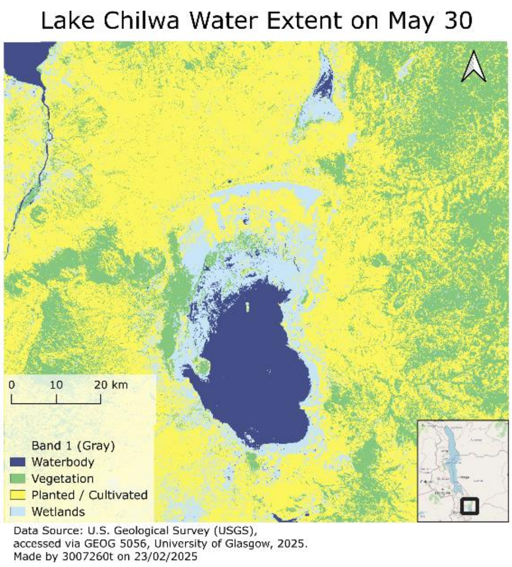


Figure 1 Flow chart of the classification task



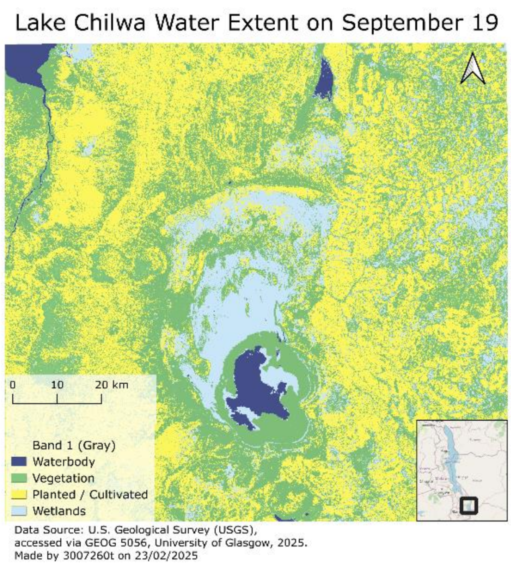


Figure 2(a)-left Classification results of May 30 and Figure 3(b)-right Classification results of September 19

Name	Reference Data						
30/05/2018 Random Forest Classifier		Waterbody	Wetlands	Vegetation	Planted / Cultivated	Total	User's Accuracy (Recall, %)
	Waterbody	312360300	0	0	5289300	317649600	98.33%
	Wetlands	2700	91415700	1758600	1138500	94315500	96.93%
	Vegetation	0	12445200	374137200	12193200	398775600	93.82%
	Planted / Cultivated	3191400	1139400	5506200	50994900	60831900	83.83%
	Total	315554400	105000300	381402000	69615900	871572600	100.00%
	Producer's Accuracy (Precision, %)	98.99%	87.06%	98.10%	73.25%	100.00%	95.10%
	Overall Accuracy	95.10%					
30/05/2018 Suppport Vector Machine Classifier		Waterbody	Wetlands	Vegetation	Planted / Cultivated	Total	User's Accuracy (Recall, %)
	Waterbody	313156800	1800	0	4491000	317649600	98.59%
	Wetlands	11700	90491400	2563200	1249200	94315500	95.95%
	Vegetation	13500	10111500	378055800	10594800	398775600	94.80%
	Planted / Cultivated	3589200	902700	5544000	50796000	60831900	83.50%
	Total	316771200	101507400	386163000	67131000	871572600	100.00%
	Producer's Accuracy (Precision, %)	98.86%	89.15%	97.90%	75.67%	100.00%	95.52%
	Overall Accuracy	95.52%					
19/09/2018 Random Forest Classifier		Waterbody	Wetlands	Vegetation	Planted / Cultivated	Total	User's Accuracy (Recall, %)
	Waterbody	57913200	4819500	0	0	62732700	92.32%
	Wetlands	0	18845100	134100	0	18979200	99.29%
	Vegetation	0	88200	48914100	233100	49235400	99.35%
	Planted / Cultivated	0	0	18000	1343700	1361700	98.68%
	Total	57913200	23752800	49066200	1576800	132309000	100.00%
	Producer's Accuracy (Precision, %)	100.00%	79.34%	99.69%	85.22%	100.00%	96.00%
	Overall Accuracy	96.00%					
19/09/2018 Suppport Vector Machine Classifier		Waterbody	Wetlands	Vegetation	Planted / Cultivated	Total	User's Accuracy (Recall, %)
	Waterbody	62720100	12600	0	0	62732700	99.98%
	Wetlands	20700	54952200	31500	0	55004400	99.91%
	Vegetation	900	100800	48899700	234000	49235400	99.32%
	Planted / Cultivated	0	0	8100	1353600	1361700	99.41%
	Total	62741700	55065600	48939300	1587600	168334200	100.00%
	Producer's Accuracy (Precision, %)	99.97%	99.79%	99.92%	85.26%	100.00%	99.76%
	Overall Accuracy	99.76%					

Table 1 Confusion Matrix of Classification by RF and SVM

4. Limitations and Improvements

1) Data Limitations:
   a) Spatial Resolution: The 30-meter resolution of Landsat 8 leads to inaccurate manual sampling. Higher resolution data (such as Sentinel-2) can improve the results.
   b) Temporal Resolution: The images in May and September are quite different. Applying multi-temporal data will better capture the seasonal dynamics of water bodies[5].
2) Processing Limitations:
   a) Atmospheric Correction: Although using Level-2 data, changing atmospheric conditions may affect quality. Specific correction could improve accuracy.
   b) Training Data: Relying on secondary sources (e.g., Google Earth) may introduce errors. Field surveys or higher-resolution references are expected in the future.
3) Classification Challenges:
   a) Class Granularity: The rough scheme (four classes) is not representative enough. A more subdivided scheme can increase the level of detail.
   b) Water-Wetland Separation: Water edges are sometimes misclassified as wetlands due to heavy reliance on NDWI.

5. Conclusion

In this study, the classification of Chilwa Lake and its surrounding areas was undertaken using Landsat-8 imagery. Two maps of classification results under Support Vector Machine algorithm were created, which highlighted substantial water changes in the lake in 2018. The writer employed the Random Forest and SVM classification algorithm to train the model, and the classification results of both algorithms were excellent, with the SVM data demonstrating superior performance (>95% overall, >98% for water). Future work would consider higher-resolution data and different algorithms to improve accuracy.

6. References

[1] L.-M. Rebelo, M. P. McCartney, and M. C. Finlayson, ‘The application of geospatial analyses to support an integrated study into the ecological character and sustainable use of Lake Chilwa’, *J. Gt. Lakes Res.*, vol. 37, pp. 83–92, Jan. 2011, doi: 10.1016/j.jglr.2010.05.004.

[2] ‘Remote Sensing Env_24-25: Assignment - Classification of Lake Chilwa, Malawi | Moodle’. Accessed: Feb. 23, 2025. [Online]. Available: https://moodle.gla.ac.uk/mod/assign/view.php?id=4557693

[3] ‘USGS EROS Archive - Landsat Archives - Landsat 8-9 OLI/TIRS Collection 2 Level-2 Science Products | U.S. Geological Survey’. Accessed: Feb. 23, 2025. [Online]. Available: https://www.usgs.gov/centers/eros/science/usgs-eros-archive-landsat-archives-landsat-8-9-olitirs-collection-2-level-2

[4] S. Shetty, P. K. Gupta, M. Belgiu, and S. K. Srivastav, ‘Assessing the Effect of Training Sampling Design on the Performance of Machine Learning Classifiers for Land Cover Mapping Using Multi-Temporal Remote Sensing Data and Google Earth Engine’, *Remote Sens.*, vol. 13, no. 8, Art. no. 8, Jan. 2021, doi: 10.3390/rs13081433.

[5] D. Yamazaki, M. A. Trigg, and D. Ikeshima, ‘Development of a global ~90m water body map using multi-temporal Landsat images’, *Remote Sens. Environ.*, vol. 171, pp. 337–351, Dec. 2015, doi: 10.1016/j.rse.2015.10.014.