ASSESSMENT OF IMAGE CLASSIFICATION ALGORITHMS FOR LAND COVER CLASSIFICATIONS IN TULLY, NY

Abishek Poudel [a, *], Shumaila Bhatti [b], Eddie Bevilacqua[a]

[a] Department of Sustainable Resources Management, SUNY ESF, Syracuse, NY 13210 USA – apoudel@syr.edu
[b] Graduate Program in Environmental Science, SUNY ESF, Syracuse, NY 13210 USA

KEYWORDS: PlanetScope, Landsat, Remote sensing, Landcover classification, ArcGIS Pro.

ABSTRACT:

The identification, delineation, and mapping of landcover is integral for resource management and planning as it establishes a baseline for thematic mapping and change detection analysis. The availability of high-resolution satellite imagery and the development of machine learning algorithms have significantly improved the prediction and accuracy of landcover classification. In this study, landcover classification is performed on seven-band Landsat 9 imagery and eight-band PlanetScope imagery for the village of Tully, NY, with an area of 900 square kilometers. The resolution of Landsat imagery is 30 meters, whereas the resolution of PlanetScope imagery is 3 meters. Classification schema is developed in ArcGIS Pro with five classification levels: conifer forest, hardwood forest, agriculture, developed, and water. Pixel-based supervised classification is performed using Support Vector Machine (SVM), Random Tress (RT), K-Nearest Neighbor (K-NN), and Maximum Likelihood Classifier (MLC). The reference dataset is acquired by an image interpreter using high-resolution imagery for map accuracy assessment. All the classification methods for Landsat imagery have more than 78% accuracy, but SVM performed best with 82% accuracy. For PlanetScope imagery, SVM performed best with 85% accuracy, whereas MLC had the lowest accuracy of 77%.

1. INTRODUCTION

1.1 Landcover Classification Using Remote Sensing

Landcover classification using remote sensing provides crucial information for environmental and natural resources monitoring, management, and change detection (Datta et al., 2022; Harezlak et al., 2020; Ullah et al., 2022). It supports informed decision-making processes and helps address various socio-environmental challenges.

By leveraging high-resolution satellite imagery and machine learning algorithms, landcover classification achieves higher accuracy, improved spatial detail, automated feature extraction, and scalability. This integration enhances the ability to monitor landcover changes, support decision-making processes, and address various environmental challenges effectively (Qing & Liu, 2022).

1.2 Satellite Choices: PlanetScope vs Landsat 9

PlanetScope is a constellation of small Earth observation satellites (typically weighing around 4 kilograms) operated by Planet, an Earth imaging company. These satellites are designed to capture high-resolution imagery of the Earth's surface daily. With its extensive coverage and frequent revisits, PlanetScope offers near-real-time monitoring and supports various applications in environmental monitoring, land management, disaster response, and more (Wang et al., 2022). PlanetScope Imagery is and eight band imagery with 3-meter resolution.

On the other hand, Landsat9 is a seven-band imagery with 30-meter resolution. Landsat 9 is the latest satellite in the Landsat program, which is a series of Earth observation satellites jointly managed by NASA and the United States Geological Survey (USGS). It follows the same repeat cycle as its predecessor, Landsat 8, capturing images of the entire Earth every 16 days (Saralioglu & Vatandaslar, 2022). This frequent revisit time provides the opportunity to monitor dynamic processes and changes happening on the Earth's surface and ensures the continuity of the Landsat program, which has been collecting imagery since 1972.

1.3 Image Classification Algorithms

In remote sensing, several image classification methods have been utilized by researchers to classify landcover and extract valuable information from imagery. In this paper we consider the following methods commonly employed:

Support Vector Machine (SVM): SVM is a supervised machine learning algorithm that has been widely used for image classification in remote sensing. It works by creating an optimal hyperplane to separate different classes in the feature space. SVM is effective in handling complex and nonlinear relationships between spectral features and landcover classes (Ahmed et al., 2018; Alimjan et al., 2017; Huang et al., 2002; Qian et al., 2015; Rana & Venkata Suryanarayana, 2020; Saralioglu & Vatandaslar, 2022; Shang et al., 2019; Waske & Benediktsson, 2007).

Random Trees (RT): RT, also known as Random Forest, is an ensemble learning algorithm that combines multiple decision trees to perform classification. Each decision tree is built on a random subset of the training data and features, and the final classification is determined by a voting mechanism. RT is known for its ability to handle high-dimensional data and provide robust classification results (Rana & Venkata Suryanarayana, 2020; Shang et al., 2019; Talukdar et al., 2020; Thanh Noi & Kappas 2018; Zafari et al., 2020).

K-Nearest Neighbor (K-NN): K-NN is a simple and intuitive classification algorithm that assigns a class label to a pixel based on the labels of its nearest neighbors in the feature space. The value of K determines the number of neighbors considered for classification. K-NN is effective when the spatial arrangement of classes is important for classification (Alimjan et al., 2017; Qian et al., 2015; Saralioglu & Vatandaslar, 2022; Thanh Noi & Kappas 2018).

Maximum Likelihood Classifier (MLC): MLC is a probabilistic classifier that assigns class labels to pixels based on the maximum likelihood estimation of the spectral distribution for each class. It assumes that the spectral values within a class follow a multivariate normal distribution. MLC is widely used due to its simplicity and robustness in handling mixed pixels (Ahmed et al., 2018; Huang et al., 2002; Waske & Benediktsson, 2007; Wang et al., 2022).

These classification methods offer different approaches to analyzing remote sensing imagery and have varying strengths and weaknesses depending on the characteristics of the data and the specific classification task at hand. Researchers often choose the most appropriate method based on the nature of the study area, the available training data, and the desired accuracy and computational efficiency requirements. The objective of this study was to evaluate the accuracy of image classification algorithms for land cover classification using Landsat 9 and PlanetScope imagery for a village namely Tully, New York.

2. METHODS

2.1 Data and Study Area

We utilized Landsat 9 and PlanetScope imagery collected on July 11, 2022 (Figure 1 and 2), to conduct our analysis. The study site was situated in the village of Tully (Figure 3), located in Central New York. The site encompassed an area of 900 square kilometers and featured diverse landcover types, including vernal pools, rolling hills, wooded forests, and agricultural fields. Tully is characterized by its rural setting and has been previously documented in studies by Kappel & Miller (2003), Kappel (2014), and Smith (2014).

Figure 1: Landsat 9, seven-band imagery with 30-meter resolution

Figure 2: PlanetScope, eight-band imagery with 3-meter resolution

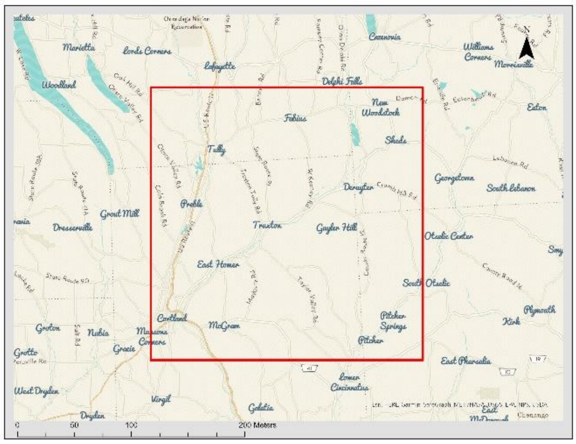

Figure 3: Study Area

2.2 Image Classification and Training Dataset Creation

To perform the landcover classification, we employed ArcGIS software. Within ArcGIS, we utilized the available tools to create a classification schema comprising five landcover categories: conifer forest, hardwood, agriculture, developed, and water. This schema was designed to encompass the major landcover types present in the study area (Figure 4).

For training dataset creation, we utilized supervised image classification techniques. The Image Analyst tool available in ArcGIS Pro facilitated this process. By leveraging the software's capabilities, we delineated representative training samples for each landcover category using the imagery data. The training samples were selected strategically to ensure a comprehensive representation of the spectral characteristics and spatial distribution within each class.

2.3 Supervised Image Classification

To achieve landcover classification, we applied various supervised image classification algorithms available within the Image Analyst tool of ArcGIS Pro. Specifically, we utilized Support Vector Machine (SVM), Random Trees (RT), K-Nearest Neighbor (KNN), and Maximum Likelihood Classification

(MLC) algorithms. These algorithms harnessed the spectral information extracted from the imagery and the corresponding training dataset.

Each algorithm was configured with carefully chosen parameters to optimize the classification performance. The input data consisted of the pre-processed Landsat 9 and PlanetScope imagery.

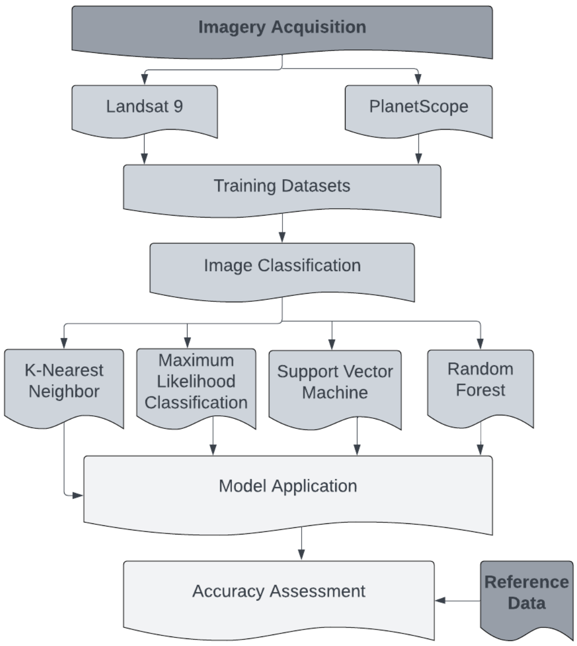

Figure 4: Workflow on landcover classification using Landsat 9 and PlanetScope imagery.

2.4 Collection of Reference Dataset

To assess the accuracy of the image classification results, we collected an independent reference dataset. This dataset was created through manual interpretation of high-resolution orthoimagery by an experienced interpreter. The interpreter employed ArcGIS tools to accurately delineate the landcover categories within the study area. The reference dataset was thoroughly validated, and ancillary data, such as ground truth measurements and existing land cover maps, were utilized to ensure its reliability.

2.5 Accuracy Assessment

To evaluate the accuracy of the image classification results, we employed several assessment techniques. A comparison was made between the classified imagery and the independently collected reference dataset. Metrics such as user's accuracy, producer's accuracy, overall accuracy, and kappa coefficient were calculated to quantitatively measure the agreement between the classified results and the reference data.

In summary, our methodology involved using ArcGIS software for image classification and training dataset creation. We applied supervised image classification algorithms (SVM, RT, KNN, and MLC) available within the Image Analyst tool of ArcGIS Pro. The reference dataset was independently collected through manual interpretation of high-resolution orthoimagery. These methods allowed us to assess the accuracy of the image classification algorithms for landcover classification in the study area.

3. RESULTS

In this study we aimed to utilize Landsat 9 and PlanetScope datasets to assess image classification algorithms for land cover classification. Table 1 summarizes the overall accuracy of image classification algorithms for land cover classification in study area. Additionally, classified map is produced for all four image classification algorithms for Landsat 9 and PlanetScope imagery respectively.

All the classification methods applied to Landsat imagery achieved an accuracy of over 78%. However, SVM exhibited the highest accuracy of 82%, outperforming the other methods. When classifying PlanetScope imagery, SVM demonstrated the highest accuracy of 85%, indicating its superior performance compared to the other methods. On the other hand, MLC showed the lowest accuracy of 77%.

In summary, for Landsat imagery, all the classification methods achieved a high accuracy of more than 78%, with SVM performing the best at 82% in line with other studies. For PlanetScope imagery, SVM outperformed the other methods with an accuracy of 85%, while MLC exhibited the lowest accuracy of 77%. SVM outperformed for both satellite imageries in line with other studies (Ahmed et al., 2018; Alimjan et al., 2017; Huang et al., 2002; Qian et al., 2015; Rana & Venkata Suryanarayana, 2020; Saralioglu & Vatandaslar, 2022; Shang et al., 2019; Waske & Benediktsson, 2007).

Table 1: Table showing overall accuracy of land cover classification for Landsat 9 and PlanetScope imagery.

Image Classification Algorithms	Overall Accuracy	
	Landsat	PlanetScope
SVM	0.82	0.85
RT	0.78	0.82
KNN	0.78	0.83
MLC	078	0.77

We found that using the SVM classification (Figure 5), water was successfully classified in both datasets due to its distinctive spectral signature, which sets it apart from other landcover categories. However, the accuracy of classifying agriculture was relatively low, with a user's accuracy of 76%. Similarly, in the Random Trees Classification, water and hardwood forest are accurately classified, while agriculture exhibits the lowest accuracy for Landsat and Planetscope datasets. These results align with the SVM classification, where agriculture also has the least accurate classification. In the K-Nearest Neighbors (KNN) classification, water was accurately classified, followed by hardwood forest. On the other hand, agriculture has the least accurate classification.

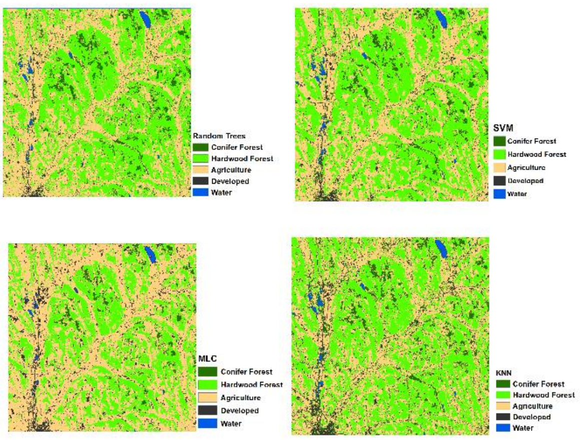

Figure 5: Classified images using Landsat 9 imagery for Random Forest, SVM, MLC and KNN algorithms.

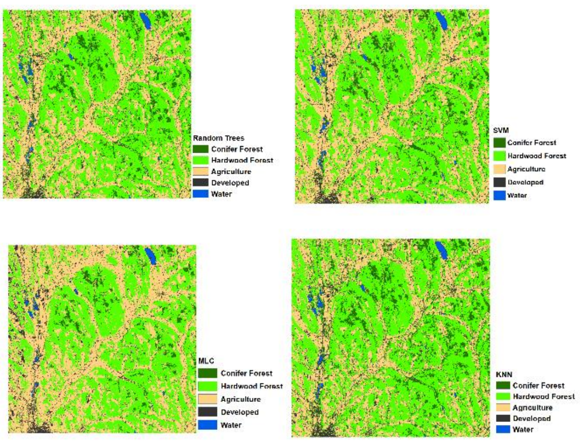

Figure 6: Classified images using PlanetScope imagery for Random Forest, SVM, MLC and KNN algorithms.

4. CONCLUSION

The image classification procedures used in this study were time-consuming due to the substantial time required for downloading and processing the imagery. This limitation highlights the need for efficient data handling and processing techniques to streamline future classification workflows.

The current results demonstrate the potential of remote sensing imagery, such as PlanetScope and Landsat 9, for pixel-based image classification. However, it is important to note that incorporating additional data sources, such as ancillary data or higher-resolution imagery, could potentially enhance the accuracy and performance of the image classification algorithms. Future studies should explore the integration of such data sources to further improve the classification results.

Furthermore, the use of field-based reference data can provide valuable validation and increase the accuracy of the generated maps. While this study relied on remote sensing imagery for classification, ground truth data collected in the field can help validate the classification results and improve the overall accuracy of the maps. The inclusion of field-based reference data should be considered in future studies to enhance the reliability and robustness of the image classification outcomes.Type text single-spaced, with one blank line between paragraphs and following headings. Start paragraphs flush with left margin.

5. REFERENCES

Ahmad, A., Kalsom, U., Mohd, O., Mawardy, M., Sakidin, H., Wahid, A., Firdaus, S., 2018. Comparative Analysis of Support Vector Machine, Maximum Likelihood and Neural Network Classification on Multispectral Remote Sensing Data. ijacsa 9. https://doi.org/10.14569/IJACSA.2018.090966

Alimjan, G., Sun, T., Jumahun, H., Guan, Y., Zhou, W., Sun, H., 2017. A Hybrid Classification Approach Based on Support Vector Machine and K-Nearest Neighbor for Remote Sensing Data. Int. J. Patt. Recogn. Artif. Intell. 31, 1750034. https://doi.org/10.1142/S0218001417500343

Datta, D., Mallick, P.K., Bhoi, A.K., Ijaz, M.F., Shafi, J., Choi, J., 2022. Hyperspectral Image Classification: Potentials, Challenges, and Future Directions. Computational Intelligence and Neuroscience 2022, e3854635. https://doi.org/10.1155/2022/3854635

Hu, B., Xu, Y., Huang, X., Cheng, Q., Ding, Q., Bai, L., Li, Y., 2021. Improving Urban Land Cover Classification with Combined Use of Sentinel-2 and Sentinel-1 Imagery. ISPRS International Journal of Geo-Information 10, 533. https://doi.org/10.3390/ijgi10080533

Huang, C., Davis, L.S., Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification. International Journal of Remote Sensing 23, 725–749. https://doi.org/10.1080/01431160110040323

Kappel, W.M., 2014. Dating Hydrologic and Geomorphic Change Using Dendrochronology in Tully Valley, Central New York: A Summary. trre 70, 91–99. https://doi.org/10.3959/1536-1098-70.2.91

Kappel, W.M., Miller, T.S., 2003. Hydrogeology of the Tully Trough, southern Onondaga County and northern Cortland County, New York (USGS Numbered Series No. 2003–4112), Hydrogeology of the Tully Trough, southern Onondaga County and northern Cortland County, New York, Water-Resources Investigations Report. U.S. Geological Survey, Reston, VA. https://doi.org/10.3133/wri034112

Planet, 2023. Planet | Homepage [WWW Document]. Planet. URL https://www.planet.com/?utm_source=google&utm_medium=paid-search&utm_campaign=discovery&utm_content=main-brd-responsive-search&gad=1 (accessed 6.16.23).

Qian, Y., Zhou, W., Yan, J., Li, W., Han, L., 2015. Comparing Machine Learning Classifiers for Object-Based Land Cover Classification Using Very High Resolution Imagery. Remote Sensing 7, 153–168. https://doi.org/10.3390/rs70100153

Qin, R., Liu, T., 2022. A Review of Landcover Classification with Very-High Resolution Remotely Sensed Optical Images—Analysis Unit, Model Scalability and Transferability. Remote Sensing 14, 646. https://doi.org/10.3390/rs14030646

Rana, V.K., Venkata Suryanarayana, T.M., 2020. Performance evaluation of MLE, RF and SVM classification algorithms for watershed scale land use/land cover mapping using sentinel 2 bands. Remote Sensing Applications: Society and Environment 19, 100351. https://doi.org/10.1016/j.rsase.2020.100351

Saralioglu, E., Vatandaslar, C., 2022. Land use/land cover classification with Landsat-8 and Landsat-9 satellite images: a comparative analysis between forest- and agriculture-dominated landscapes using different machine learning methods. Acta Geod Geophys 57, 695–716. https://doi.org/10.1007/s40328-022-00400-9

Shang, M., Wang, S., Zhou, Y., Du, C., Liu, W., 2019. Object-based image analysis of suburban landscapes using Landsat-8 imagery. International Journal of Digital Earth 12, 720–736. https://doi.org/10.1080/17538947.2018.1474959

Smith, R.G., 2014. Factors affecting primary production and respiration in small forested pools Heiberg Forest, Tully, NY (Honors Theses. Paper 31.).

Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal, S., Liou, Y.-A., Rahman, A., 2020. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. Remote Sensing 12, 1135. https://doi.org/10.3390/rs12071135

Thanh Noi, P., Kappas, M., 2018. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors 18, 18. https://doi.org/10.3390/s18010018

Ullah, I., Aslam, B., Shah, S.H.I.A., Tariq, A., Qin, S., Majeed, M., Havenith, H.-B., 2022. An Integrated Approach of Machine Learning, Remote Sensing, and GIS Data for the Landslide Susceptibility Mapping. Land 11, 1265. https://doi.org/10.3390/land11081265

Wang, J., Yang, M., Chen, Z., Lu, J., Zhang, L., 2022. An MLC and U-Net Integrated Method for Land Use/Land Cover Change Detection Based on Time Series NDVI-Composed Image from PlanetScope Satellite. Water 14, 3363. https://doi.org/10.3390/w14213363

Waske, B., Benediktsson, Jó.A., 2007. Fusion of Support Vector Machines for Classification of Multisensor Data. IEEE Transactions on Geoscience and Remote Sensing 45, 3858–3866. https://doi.org/10.1109/TGRS.2007.898446

Yang, X., 2011. Parameterizing Support Vector Machines for Land Cover Classification. photogramm eng remote sensing 77, 27–37. https://doi.org/10.14358/PERS.77.1.27

Zafari, A., Zurita-Milla, R., Izquierdo-Verdiguier, E., 2020. A Multiscale Random Forest Kernel for Land Cover Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13, 2842–2852. https://doi.org/10.1109/JSTARS.2020.2976631