

PEMANFAATAN ALGORITMA RANDOM FOREST UNTUK KLASIFIKASI TUTUPAN LAHAN DI KABUPATEN KAPUAS DAN PULANG PISAU

ANANDA KURNIA, INDRIANAWATI

Institut Teknologi Nasional Bandung
Email: kisekiou14@gmail.com

ABSTRAK

Pengembangan metodologi baru untuk memetakan tutupan lahan telah menjadi salah satu perhatian utama pada skala regional dan global. Tutupan lahan harus diprioritaskan untuk dipetakan secara cepat sebab menjadi parameter penting dalam banyak aplikasi pekerjaan. Penelitian ini bertujuan untuk menerapkan salah satu algoritma klasifikasi, yaitu random forest, untuk klasifikasi tutupan lahan di Kabupaten Kapuas dan Pulang Pisau dengan parameter atau data yang digunakan untuk klasifikasi tutupan lahan yaitu ketinggian, kemiringan, dan 10 band dari citra satelit Sentinel-L2A. Model yang dibuat memiliki ntree atau jumlah pohon yang akan ditumbuhkan sebanyak 300 pohon acak dan mtry atau jumlah parameter yang akan digunakan adalah 4 parameter acak. Dari hasil klasifikasi diperoleh kelas tutupan lahan hutan menjadi kelas dengan luasan terbesar dan kelas tutupan lahan perairan menjadi kelas dengan luasan terkecil dibandingkan dengan 4 kelas lainnya. Namun, berdasarkan akurasi validation dataset, model 300 pohon acak dan 4 parameter acak ini memiliki akurasi yang kurang baik (overfit) sehingga belum direkomendasikan untuk digunakan membuat peta tutupan lahan.

Kata kunci: tutupan lahan, random forest, machine learning, ntree

1. PENDAHULUAN

Pengembangan metodologi baru untuk memetakan tutupan lahan dengan menggunakan citra penginderaan jauh telah menjadi salah satu perhatian utama pada skala regional dan global (Singh dkk., 2021). Cara paling signifikan untuk memproduksi peta tutupan lahan adalah klasifikasi dengan citra satelit menggunakan beragam algoritma klasifikasi (Avci, 2023). Tutupan lahan harus diprioritaskan untuk dipetakan dengan cepat, karena tutupan lahan menjadi parameter penting dalam banyak aplikasi pekerjaan (Mathieu dkk., 2007; Shi dkk., 2015; Ozdarici-Ok dkk., 2015; Zhang dan Kovacs, 2012; Ardila dkk., 2011; Fauvel dkk., 2013). Salah satu kegunaan dari tutupan lahan ini adalah dapat digunakan oleh pemerintah untuk merumuskan kebijakan dan strategi pembangunan berkelanjutan (Giri, 2012).

Terdapat beberapa metode yang umum digunakan untuk memetakan tutupan lahan, baik itu manual ataupun semi-otomatis. Metode pemetaan tutupan lahan secara manual dapat dilakukan dengan *on-screen digitizing*, namun metode ini rentan bersifat subjektif. Artinya adalah pandangan atau persepsi setiap orang dapat berbeda-beda ketika mendefinisikan batas antar kelas tutupan lahan, mengakibatkan hasil digitasi yang berbeda untuk setiap orang apabila dikerjakan secara manual (Wijaya, 2021). Untuk mengatasi subjektivitas tersebut, banyak penelitian yang memanfaatkan data penginderaan jauh dan berbagai algoritma klasifikasi *machine learning* untuk melakukan pemetaan tutupan lahan yang lebih cepat,

murah, dan efisien (Allam dkk, 2019; El-Zeiny & Effat, 2017; Mandanici & Bitelli, 2015; Mohamed & El-Raey, 2019).

Dalam pemetaan, *machine learning* telah banyak diterapkan untuk melakukan klasifikasi tutupan lahan. Yimer (2024) mengkaji algoritma *machine learning* mana yang memiliki performa paling baik untuk klasifikasi tutupan lahan, adapun algoritma yang dikaji yaitu *random forest* (RF), *support vector machine* (SVM), *maximum likelihood classification* (MLC), *classification & regression tree* (CART), *K-nearest neighbor* (KNN), dan *artificial neural network* (ANN), hasil kajian menunjukkan RF menduduki posisi pertama, lalu disusul algoritma lainnya. Avci (2023) mengkaji algoritma *machine learning* yang memiliki performa paling baik dalam mengklasifikasi tutupan lahan yakni *random forest* dan *support vector machine* (SVM), hasil kajian menunjukkan bahwa algoritma *random forest* merupakan algoritma dengan performa paling baik untuk klasifikasi tutupan lahan. Nguyen (2018) menggunakan klasifikasi *random forest* untuk membuat peta tutupan lahan dari citra Landsat 8 OLI, dengan membuat 4 model dengan *mtry* dan *ntree* berbeda pada tiap modelnya untuk melihat apakah terdapat keterkaitan antara jumlah *mtry* dan *ntree* yang digunakan terhadap akurasi. Dari beberapa penelitian yang telah dilakukan sebelumnya, dapat ditarik benang merah bahwa algoritma *random forest* sangat mumpuni dalam klasifikasi tutupan lahan terbukti dari hasil komparasinya dengan algoritma *machine learning* yang lain ataupun penelitian yang pernah menggunakan *random forest* untuk klasifikasi tutupan lahan.

Berdasarkan uraian tersebut, dalam penelitian ini akan mencoba menerapkan algoritma *random forest* untuk klasifikasi tutupan lahan di Kabupaten Kapuas dan Pulang Pisau. Parameter atau data yang digunakan untuk klasifikasi tutupan lahan meliputi ketinggian, kemiringan, dan 10 band dari citra satelit Sentinel-L2A. Adapun jumlah *ntree* atau pohon yang akan ditumbuhkan sebanyak 300 pohon acak dan *mtry* atau jumlah parameter yang digunakan sebanyak 4 parameter acak. Batasan masalah dalam penelitian ini adalah (1) kelas tutupan lahan yang dibuat terdiri dari sawah, perkebunan, lahan terbangun, hutan, dan perairan; (2) rasio *training dataset* sebesar 70% dan *validation dataset* 30% (Thien dan Yeo, 2021); (3) parameter yang digunakan untuk proses klasifikasi tutupan lahan adalah band dari citra satelit Sentinel-2 yaitu: B2 biru, B3 hijau, B4 merah, B5–B7, B8 sebagai band Near Infra Red (NIR), B8A Vegetation red edge, dan B11-B12 band Short Wave Infra Red (SWIR) (Svoboda dkk, 2022).

2. DATA DAN METODOLOGI PENELITIAN

2.1 Data Penelitian

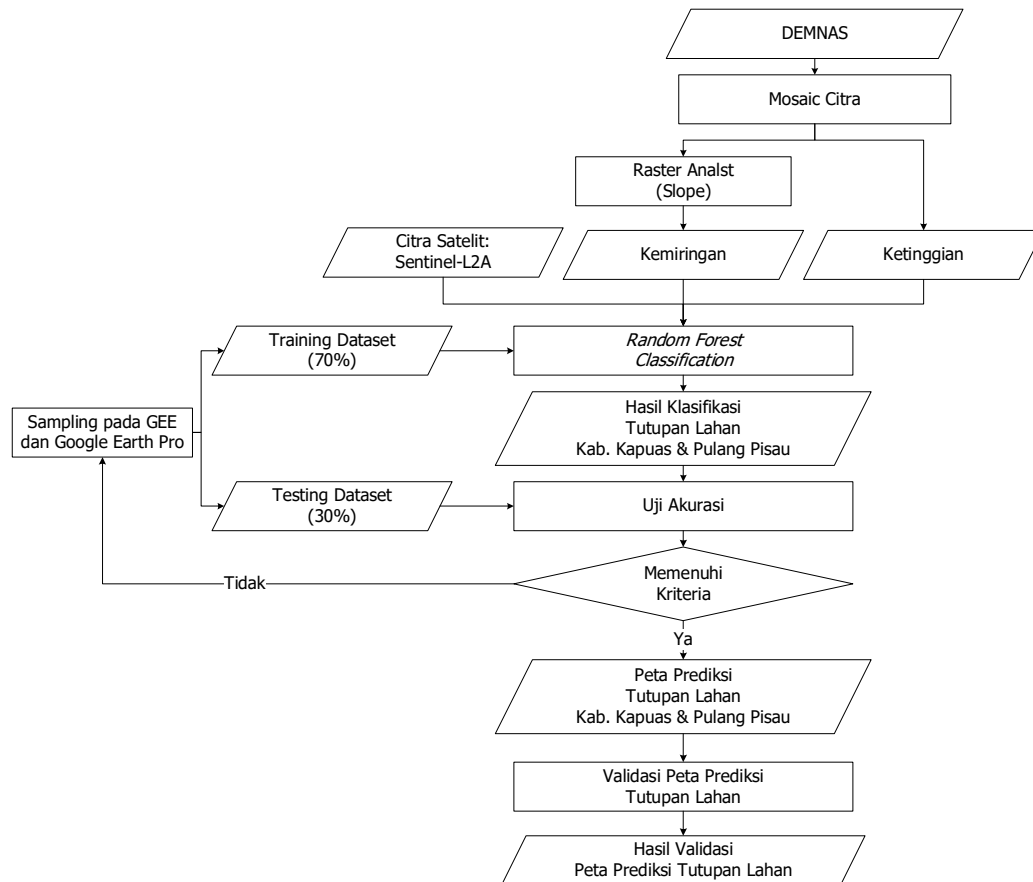
Data yang digunakan dalam penelitian ini dirincikan pada Tabel 1.

Tabel 1. Data Penelitian

No.	Jenis Data	Resolusi	Tahun	Format Data	Sumber Perolehan Data
1.	Citra Satelit Sentinel-L2A	10 & 20 meter	2023	.TIFF	Copernicus
2.	Data Ketinggian (DEMNAS)	0,27 Arcsecond	2023	.TIFF	Badan Informasi Geospasial (BIG)
3.	Data Kemiringan (DEMNAS)	0,27 Arcsecond	2023	.TIFF	Badan Informasi Geospasial (BIG)
4.	Administrasi Kabupaten Kapuas dan Pulang Pisau	1:50.000	2023	.shp	Badan Informasi Geospasial (BIG)

2.2 Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini ditunjukkan pada Gambar 1.



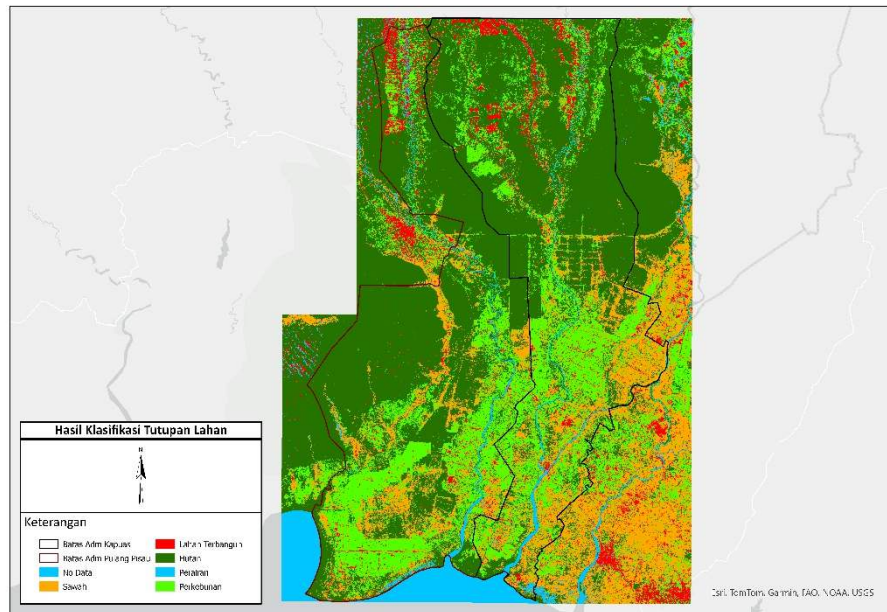
Gambar 1. Diagram Alir Penelitian

Secara umum tahapan penelitian dibagi dalam beberapa tahap, yaitu:

- 1) Pengumpulan data seperti data DEM Nasional (DEMNAS), dan data citra satelit Sentinel-L2A yang bersumber dari Copernicus.
- 2) Pengolahan data DEMNAS dengan menggunakan perangkat lunak ArcGIS Pro. Data DEMNAS dimosaik terlebih dahulu dan menghasilkan data ketinggian dalam bentuk raster. Selanjutnya data ketinggian diolah kembali menggunakan *tools raster analyst (slope)* hingga menghasilkan data kemiringan dalam bentuk raster.
- 3) Pemilihan titik sampel pada wilayah studi dengan bantuan interpretasi dari *platform* Google Earth Engine (GEE) dan Google Earth Pro. Titik sampel yang dikeluarkan berformat *shapefile (.shp)*.
- 4) Klasifikasi tutupan lahan dengan algoritma *random forest* pada GEE.
- 5) Evaluasi hasil prediksi tutupan lahan yang telah dilakukan. Jika tutupan lahan memenuhi akurasi yang telah disepakati, maka dapat dilanjutkan hingga menjadi peta tutupan lahan. Namun jika akurasi tidak memenuhi, maka proses pemilihan titik sampel harus dilaksanakan kembali.

3. HASIL DAN PEMBAHASAN

Proses klasifikasi tutupan lahan menggunakan algoritma *random forest* dilakukan dengan membuat model dimana *ntree* atau jumlah pohon yang ditumbuhkan sebanyak 300 pohon acak dan *mtry* atau jumlah parameter yang digunakan sebanyak 4 parameter acak. Sesuai dengan proses klasifikasi tutupan lahan dengan algoritma *random forest* yang dilakukan pada platform Google Earth Engine, hasil klasifikasi tutupan lahan dengan 300 pohon acak dan 4 parameter acak ditunjukkan pada Gambar 2.



Gambar 2. Hasil Klasifikasi Tutupan Lahan dengan 300 Pohon Acak dan 4 Parameter Acak

Luasan area tutupan lahan (dalam satuan Ha) dari model klasifikasi *random forest* dapat dilihat pada Tabel 2.

Tabel 2. Luasan Tutupan Lahan Hasil Klasifikasi *Random Forest*

Kelas	Model (300 <i>ntree</i> & 4 <i>mtry</i>)
Sawah (1)	513.400
Lahan Terbangun (2)	173.305
Hutan (3)	1.539.043
Perairan (4)	67.673
Perkebunan (5)	629.753
Total	2.923.175

Hutan menjadi kelas tutupan lahan dengan luasan paling besar, sedangkan perairan menjadi kelas tutupan lahan dengan luasan paling kecil pada model yang dibuat.

a. Uji Akurasi Hasil Klasifikasi Tutupan Lahan

Proses analisis dilakukan dengan melihat nilai *overall accuracy* dan koefisien kappa dari setiap model yang dibangun berdasarkan *training dataset* dan *validation dataset*. Analisis juga dilakukan untuk melihat pengaruh setiap paramater dalam pembentukan model dan akurasi

berdasarkan hasil perhitungan *Mean Decrease Gini* (MDG) dan *Mean Decrease Accuracy* (MDA). *Confussion matrix* mendefinisikan kesalahan dalam klasifikasi peta tutupan lahan dapat terjadi karena adanya piksel citra pada suatu kelas yang pada dasarnya ditugaskan pada kelas lain yang mengakibatkan terjadinya *false positive* dan *false negative*. *False positive* dan *false negative* dapat terjadi ketika sebuah piksel citra diklasifikasikan secara tidak benar dan tidak termasuk ke dalam kelas tersebut. Kesalahan di satu kelas dapat mengakibatkan kesalahan komisi di kelas lain (Chugtai, 2021). *Confussion matrix* dari *training dataset*, *testing dataset*, dan *validation dataset* model 300 *ntree* dan 4 *mtry* dapat dilihat pada Tabel 3 s.d. Tabel 5.

Tabel 3. Confussion Matrix dari Training Dataset

		Predicted				
		1	2	3	4	5
Actual	1	104	0	0	0	0
	2	0	102	0	0	0
	3	0	1	116	0	0
	4	0	0	0	98	0
	5	0	0	0	0	110

Tabel 4. Confussion Matrix dari Testing Dataset

		Predicted				
		1	2	3	4	5
Actual	1	47	3	0	0	1
	2	2	48	0	1	2
	3	1	1	34	0	2
	4	1	1	0	47	0
	5	1	1	1	0	42

Tabel 5. Confussion Matrix dari Validation Dataset

		Predicted				
		1	2	3	4	5
Actual	1	5	0	10	0	10
	2	5	14	0	1	5
	3	3	1	17	0	4
	4	0	1	0	21	0
	5	8	0	3	0	14

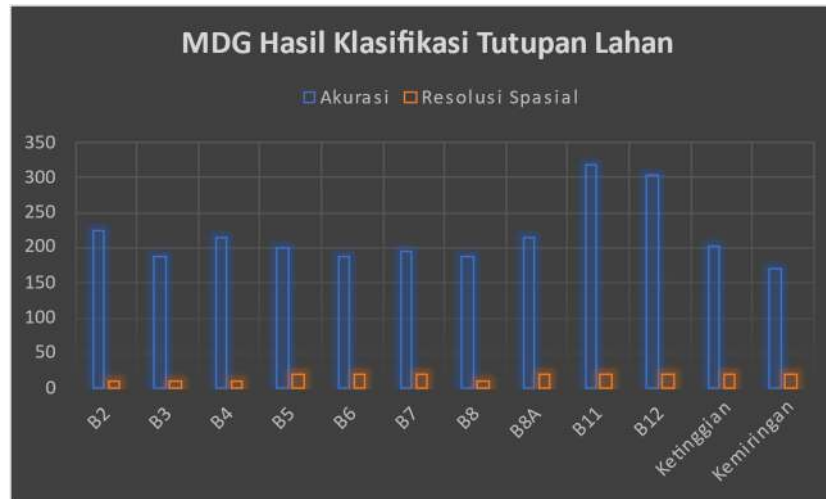
Berdasarkan hasil perhitungan nilai *overall accuracy* dan koefisien kappa dari 300 pohon acak berdasarkan *training dataset*, *testing dataset*, dan *validation dataset* didapatkan hasil sebagai seperti yang ditunjukkan pada Tabel 6.

Tabel 6. Overall Accuracy dan Koefisien Kappa dari Model yang Dihasilkan

Data	Overall Accuracy (%)	Koefisien Kappa (%)
Training Dataset	99,81%	99,76%
Testing Dataset	92,37%	90,42%
Validation Dataset	58,19%	47,71%

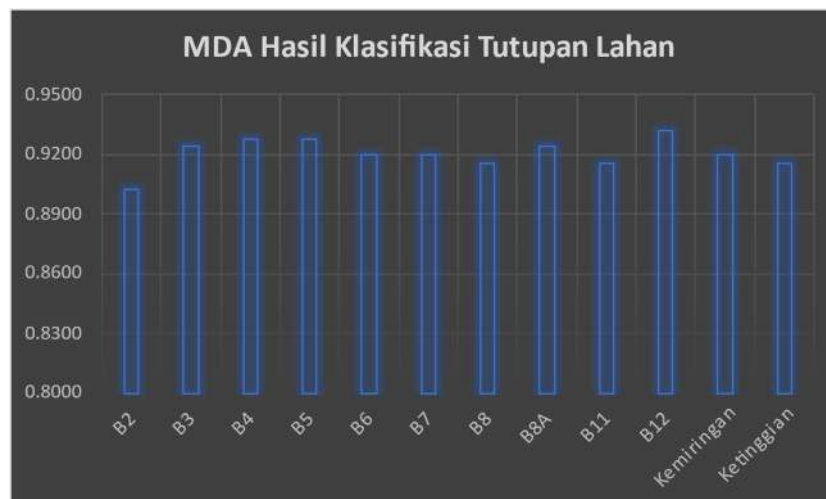
Hasil perhitungan *overall accuracy* dari ketiga data menunjukkan bahwa akurasi dari model mengalami *overfit*. Nilai *overall accuracy* yang dianggap layak oleh Anderson (1976) adalah lebih dari atau sama dengan 0,85 di mana nilai ini terpenuhi pada *training dataset* dan *testing*

dataset, namun tidak terpenuhi pada *validation dataset*. Akurasi berdasarkan nilai koefisien kappa jika mengacu pada klasifikasi yang dilakukan oleh Landis dan Koch (1977), nilai koefisien kappa dari model prediksi termasuk kelas *moderate*. Perhitungan pengaruh parameter terhadap pembentukan model berdasarkan nilai *Mean Decrease Gini* (MDG) didapatkan hasil seperti yang ditunjukkan pada Gambar 3.



Gambar 3. MDG Model-2

Berdasarkan grafik nilai MDG pada Gambar 3, maka band 11 merupakan parameter yang paling sering digunakan sebagai *decision node* pembentuk model sedangkan Kemiringan merupakan parameter yang paling sedikit digunakan sebagai *decision node* pembentuk model. Perhitungan pengaruh setiap parameter terhadap akurasi yang dihasilkan dari hasil pembentukan model berdasarkan nilai *Mean Decrease Accuracy* (MDA) didapatkan hasil seperti yang ditunjukkan pada Gambar 4.



Gambar 4. MDA Model-2

Hasil perhitungan menunjukkan bahwa seluruh parameter mempunyai pengaruh yang besar terhadap penurunan akurasi. Parameter yang paling berpengaruh terhadap penurunan akurasi adalah band 12 dengan nilai penurunan akurasi sebesar lebih dari 92%, sedangkan parameter yang mempunyai pengaruh paling kecil terhadap penurunan akurasi adalah band 2 dengan nilai penurunan akurasi lebih dari 89%.

4. KESIMPULAN

Proses klasifikasi tutupan lahan dilakukan dengan membuat model dimana *ntree* atau jumlah pohon yang akan ditumbuhkan sebanyak 300 pohon acak dan *mtry* atau jumlah parameter yang digunakan sebanyak 4 parameter acak. Berdasarkan hasil klasifikasi tutupan lahan, kelas hutan menjadi kelas tutupan lahan dengan luasan paling besar pada model yang dibuat, sedangkan perairan menjadi kelas tutupan lahan dengan luasan paling kecil. Model dengan 300 pohon acak dan 4 parameter acak memiliki akurasi yang kurang baik pada *validation dataset*. Hal ini menunjukkan model yang dibuat mengalami *overfit* secara akurasi. Oleh karena itu, model prediksi yang sudah dibuat belum dapat direkomendasikan untuk digunakan membuat peta tutupan lahan.

DAFTAR PUSTAKA

- Allam, M., Bakr, N., & Elbably, W. (2019). Multi-temporal assessment of land use/land cover change in arid region based on landsat satellite imagery: Case study in Fayoum Region, Egypt. *Remote Sensing Applications: Society and Environment*, 14, 8–19.
- Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*. US Government Printing Office (Vol. 964).
- Ardila, J. P., Tolpekin, V. A., Bijker, W., & Stein, A. (2011). Markov random field-based superresolution mapping for identification of urban trees in vhr images. *Photogrammetry Remote Sensing*, 66(6), 762–775.
- Avci, C., Budak, M., Yagmur, N., & Balcik, F. B. (2023). Comparison between random forest and support vector machine algorithms for LULC classification. *International Journal of Engineering and Geosciences*, 8(1), 1–10.
- Chughtai, A. H., Abbasi, H., & Karas, I. R. (2021). A review on change detection method and accuracy assessment for land use land cover. *Remote Sensing Applications: Society and Environment*, 22, 100482.
- El-Zeiny, A. M., & Effat, H. A. (2017). Environmental monitoring of spatiotemporal change in land use/land cover and its impact on land surface temperature in El-Fayoum governorate, Egypt. *Remote Sensing Applications: Society and Environment*, 8, 266–277.
- Fauvel, M., Tarabalka, Y., Benediktsson, J. A., Chanussot, J., & Tilton, J. C. (2013). Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3), 652–675.
- Giri, C. P. (2012). *Remote Sensing of Land Use and Land Cover Principles and Applications*. CRC Press.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33, 159–174.
- Mandanici, E., & Bitelli, G. (2015). Multi-image and multisensor change detection for long-term monitoring of arid environments with landsat series. *Remote Sensing*, 7(10), 14019–14038.
- Mathieu, R., Freeman, C., & Aryal, J. (2007). Mapping private gardens in urban areas using object-oriented techniques and very high-resolution satellite imagery. *Landscape and Urban Planning*, 81(3), 179–192.
- Mohamed, S. A., & El-Raey, M. E. (2019). Land cover classification and change detection analysis of Qaroun and Wadi El-Rayyan lakes using multi-temporal remotely sensed imagery. *Environmental Monitoring and Assessment*, 191(4).
- Nguyen, H. T. T., Doan, T. M., & Radeloff, V. (2018). Applying Random Forest classification to map Land use/Land cover using Landsat 8 OLI. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42(3W4), 363–367.

- Ozdarici-Ok, Ok, A. O., & Schindler, K. (2015). Mapping of agricultural crops from single high-resolution multispectral images data-driven smoothing vs parcel-based smoothing. *Remote Sensing*, 7(5), 5611–5638.
- Shi, H., Chen, L., Bi, F., Chen, H., & Yu, Y. (2015). Accurate Urban Area Detection in Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters*, 12(9).
- Singh, R. K., Singh, P., Drews, M., Kumar, P., Singh, H., Gupta, A. K., Govil, H., Kaur, A., & Kumar, M. (2021). A machine learning-based classification of LANDSAT images to map land use and land cover of India. *Remote Sensing Applications: Society and Environment*, 24.
- Svoboda, J., Stych, P., Lastovicka, J., Paluba, D., & Kobliuk, N. (2022). Random Forest Classification of Land Use, Land-Use Change and Forestry (LULUCF) Using Sentinel-2 Data—A Case Study of Czechia. *MDPI Journal*.
- Thien, T. F., & Yeo, W. S. (2021). A Comparative Study Between PCR, PLSR, And LWPLS On The Predictive Performance At Different Data Splitting Ratios. *Chemical Engineering Communications*.
- Wijaya, C., Yusiyaniti, I., & Piero, N. J. (2021). Deteksi Otomatis Garis Pantai Menggunakan Teknologi Machine Learning Dengan Coastsat. *In Prosiding Sriwijaya Geology Festival*.
- Yimer, S. M., Bouanani, A., Kumar, N., Tischbein, B., & Borgemeister, C. (2024). Comparison of different machine-learning algorithms for land use land cover mapping in a heterogenous landscape over the Eastern Nile river basin, Ethiopia. *Advances in Space Research*, 74(5), 2180–2199.
- Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: a review. *Precision Agriculture*, 13(6), 693–712.