第 49 卷第 2 期

七月 2025

Vol. 50 No. 1

Jan - March 2025

Diqiu Kexue - Zhongguo Dizhi Daxue Xuebao/Earth Science Journal of China University of Geosciences

https://dqkx-periodicals.com

DOI: 10.5281/zenodo. 15113567

Enhancing War on Drug Effort by Leveraging Remote Sensing Data and Machine Learning

Dedi Irawadi[1,2,a)], Tuga Mauritsius[1,b)]

[1]*Bina Nusantara University, Information Systems Management Department, BINUS Graduate Programs – Master of Information Systems Management, Bina Nusantara University, Jakarta, 11480, Indonesia*

[2]*Research Center for Satellite Technology, Research Organization for Aeronautics and Space, National Research and Innovation Agency (BRIN), Bogor, 16310, Indonesia*

[a)]Corresponding author: dedi004@brin.go.id; dedi.irawadi@binus.ac.id

[b)]Tmauritsus@binus.edu

Published: 31 March 2025

Accepted: 29 January 2025

Received: 19 December 2024

Abstract: Traditional methods of eradicating illicit cannabis plantations that relies on informants are ineffective. Integrating remote sensing data with machine learning enhances the speed and accuracy of detection, improving eradication efforts. Several studies for smart cannabis detection have been developed in other regions of the world, however, no publications are recorded to implement the utilization of combination of remote sensing data with machine learning algorithms for cannabis plantation detection in Indonesia. This study proposes a fast methodology using remote sensing and machine learning to identify potential cannabis plantations in Indonesia. A Random Forest (RF) model outperforms a Gradient Boosting Tree (GBT) in detecting cannabis plantations, achieving higher accuracy (90.18% overall, Kappa = 0.8309) than GBT (84.89% overall, Kappa = 0.7712). The RF-based model also demonstrates superior spatial performance with less misclassification. Although the proposed methodology performs well, it is limited to homogeneous cannabis plantations. Future research should also address high cloud cover, which complicates detection and increases the risk of overfitting.

Keywords: *: Remote sensing, machine learning, cannabis, smart detection, law enforcement*

1. Introduction

Indonesian law enforcement faces significant challenges in combating drug trafficking, as Cannabis Sativa L. or cannabis or marijuana accounts for 65.5% of illicit drug abuse (Putri et al., 2022) and remains widely available. Supply of cannabis mainly comes from illegal plantations that span thousands of hectares in Aceh and North Sumatra, often located in remote, forested, and hilly regions, making detection difficult. Traditional eradication methods have proven ineffective. Remote sensing satellites, which provide regular and extensive coverage, offer a more efficient monitoring solution. Combined with advancements in machine learning and the increasing affordability of satellite data, these technologies enhance the identification of cannabis plantations and support more effective law enforcement efforts.

Cannabis cultivation requires high solar intensity (Morello et al., 2022) and sufficient water (Dillis et al., 2020), it implies that the location of cannabis plantations should be aerially visible and close to water sources. And remote sensing satellites are able to capture aerially visible earth's surface data. Land-use and land-cover dynamic can be tracked and analysed using multitemporal remote sensing data (Armenteras et al., 2011; Rodriguez-Galiano & Chica-Rivas, 2014; Sirro et al., 2018). Landsat 8, Sentinel 1 and Sentinel 2 are used to predict yield of winter wheat (Fang et al., 2020; Zhang et al. 2023), pastoral information (Zhu et al., 2023), and estimation of tobacco crop (Khan et al., 2020) with a good accuracy.

Previous studies demonstrate that combining remote sensing data and machine learning can used as an effective tools to detect illegal cannabis plantations. Based on the area of eradication activity records obtained from the law enforcement agencies, datasets of cannabis plantations can be developed to train machine learning algorithms for cannabis plantations detection (Ferreira et al., 2019). Methods include the use of Normalized Vegetation Index (NDVI) from Landsat 8 (Mattiuzzi et al., 2014), high-resolution IKONOS data with deep learning (Ferreira et al., 2019), medium-resolution fusion data with machine learning (Sujud et al., 2021), and high-resolution data (SPOT 5 & QuickBird) with machine learning (Bicakli et al., 2022) yield overall accuracy greater than 80%. However, all aforenoted studies deal with mostly large area of cannabis plantations (3 to 9 Ha), posing challenges for applications in Indonesia's unique geographical characteristic and fragmented areas.

This paper proposes a methodology to detect the potential location of cannabis plantations in Indonesia, particularly in Aceh Bear, Nanggroe Aceh Darussalam (NAD), by utilizing machine learning and multitemporal medium-resolution Sentinel 2 data. Cloud-free multitemporal medium-resolution data were sorted to build the datasets. Based on the record of eradication activities, datasets of cannabis and non-cannabis were extracted. Two machine learning algorithms will be used and assessed to process the dataset to show their performance and accuracy. Potential locations of illegal cannabis plantations can be investigated further to determine the condition of cannabis plants to be used for eradication programs. This methodology offers to enhance and increase the effectiveness of program war on drug in Indonesia to eradicate the illegal cannabis plantations.

2. Material And Methods

2.1. Study Area

Figure 1 *Study Area: Nanggroe Aceh Darussalam*

The study area for this research is NAD, due to the availability of on-site data recorded in law enforcement-led cannabis eradication operations as illustrated in Figure 1. The field reports recorded the date and location of the eradication operations done by law enforcement agencies as illustrated in Table 1. The on-site reports also show that the cannabis cultivation in this region does not mix with other vegetation; it is absence of intercropping. These facts will help collection of training dan test datasets to be more accurate.

***Table 1** Report of eradication operations*

PEMUSNAHAN LADANG GANJA 2022											
No.	DATEIMAGES	Waktu Pelaksanaan	Koordinat (DMS)	Koordinat (DD)	LAT	LON	Lokasi	Luas (m²)	Jumlah Tanaman (Batang)	Berat Tanaman (Ton)	Keterangan
1.	3/12/2022	3/15/2022					Dusun Meurah Desa Lamteuba, Kecamatan Seulimeum Kabupaten Aceh Besar Provinsi Aceh	5,830	4,000	2	-
2.	3/13/2022	3/16/2022					Dusun Meurah Desa Lamteuba, Kecamatan Seulimeum Kabupaten Aceh Besar Provinsi Aceh	2,083	9,000	4.5	-
3.	5/28/2022	5/31/2022					Pegunungan Leuser Desa Agusen Kecamatan Blangkejeren Kabupaten Gayo Lues Provinsi Aceh	20,000	10,000	5	-
4.	5/28/2022	5/31/2022					Pegunungan Leuser Desa Agusen Kecamatan Blangkejeren Kabupaten Gayo Lues Provinsi Aceh	30,000	10,000	5	-
5.	6/25/2022	6/28/2022					Desa Teupin Reusep Kecamatan Sawang Kabupaten Aceh Utara Provinsi Aceh	40,000	15,000	7.5	-
6.	6/25/2022	6/28/2022					Desa Blang Manyak Kecamatan Sawang Kabupaten Aceh Utara Provinsi Aceh	20,000	5,000	2.5	-
							Desa [illegible] Kec. [illegible] Kab. [illegible]				

2.2. Remote Sensing Data

This study utilized 10-m resolution Sentinel-2 optical data from 2021 to 2023. Due to cloud cover susceptibility, a selection process was applied to obtain cloud-free imagery for cannabis dataset extraction. Datasets of cannabis and non-cannabis were generated from multi-temporal Sentinel-2 data based on records of eradication operations. Validation, essential for accurate datasets, relied on high-resolution imageries from SuperDove (PlanetScope) and other satellites constellation (Maxar and Airbus).

2.3. Methodology

The proposed detection model utilizes multi-temporal Sentinel 2 data, on-site data and high-resolution data for validation as illustrated in Figure 2. Due to high cloud cover in the area of interest, data preparation mainly dealt with data selection and cleaning. Datasets were built in 64x64 pixels in locations listed in the record of cannabis plantations eradication operations, and selected within 6 months before and 3 months after the date of the operation. All datasets were validated using available high-resolution satellite data. Supervised classification was then performed by selected machine learning algorithms to produce map of potential cannabis plantations. Further, this information can be analysed using vegetation index to determine the real-time condition of these potential plantations

Artificial Intelligence (AI), comprises algorithms that enable computers to generate their own algorithms without explicit programming, forming a system capable of learning how to learn (Corea, 2017). Corea (2017) further asserts that AI holds significant technological potential. Machine learning, a subset of AI, advances the democratization of intelligence by empowering machines to acquire knowledge from data and autonomously enhance their performance.

Figure 2 *Cannabis Detection Model Flowchart*

Leo Breiman and Adele Cutler initially created Random Forest (RF), a popular machine learning technique that combines the outputs of several decision trees to get a single prediction. Its adaptability and ease of implementation contribute to its broad application in both classification and regression tasks (*What Is Random Forest? | IBM*, n.d.). According to Breiman (2001), Random Forest (RF) is an ensemble-based classification method that improves performance by integrating multiple decision tree classifiers. Its advantages include the ability to handle both numerical and categorical features, robustness to missing data and outliers, and the provision of feature relevance assessment, facilitating model interpretability. Additionally, RF exhibits algorithmic stability, remaining relatively insensitive to parameter changes. Random Forest (RF) consists of multiple predictor trees, each trained on randomly sampled vector values drawn independently from the same distribution. As the number of trees grows, the ensemble's generalization error stabilizes. This error is influenced by both the strength of individual trees and their correlation. (Breiman, 2001).

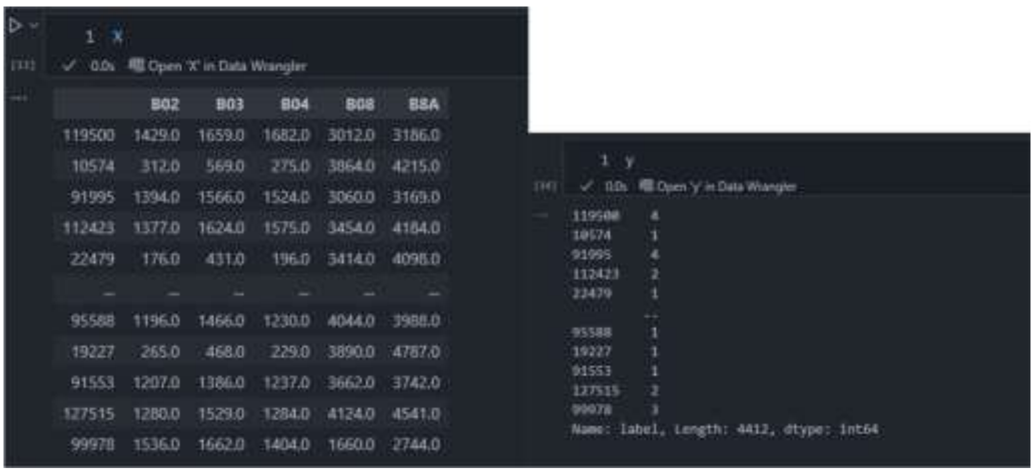
According to Friedman (2002) in Sujud et al. (2021), Gradient Boosting Tree (GBT) is a tree-based machine learning technique similar to Random Forest (RF), optimizing its algorithm by minimizing the loss function through gradient descent. GBT prioritizes shallow decision trees or weak learners, assigning higher weights to misclassified pixels to improve classification accuracy. With each additional decision tree, the total error decreases (Callens et al. (2020) in Sujud et al. (2021)). Dimov et al. (2017) reported that although GBT is more prone to overfitting than RF, it outperforms Support Vector Machines (SVM) and Multi-Layer Perceptron (MLP) in agricultural land detection accuracy.

3. Results And Discussion

Table 2 *Scene Classification Layer (SLC) adopted fromPolyGeo (2022)*

Label	Classification
0	NO_DATA
1	SATURATED_OR_DEFECTIVE
2	DARK_AREA_PIXELS
3	CLOUD_SHADOWS
4	VEGETATION
5	NOT_VEGETATED
6	WATER
7	UNCLASSIFIED
8	CLOUD_MEDIUM_PROBABILITY
9	CLOUD_HIGH_PROBABILITY
10	THIN_CIRRUS
11	SNOW

Taking advantages of Scene Classification Layer (SLC) illustrated in Table 2, Sentinel-2 data could be masking into 11 classes. By merging similar feature classes, this data could be simplified into 6 classes (no data/unclassified; dark area; high vegetation; shrub; water body; cloud) and cannabis plantation class. Sentinel 2 data was then organized into a tabular format suitable for machine learning applications as illustrated in Figure 3. To optimize processing, only 10-m resolution bands (Band 2, 3, 4, and 8) representing visible light (blue, green, red) and near-infrared are used. Additionally, the 20-m Band 8A (NIR vegetation edge) is included for its sensitivity to vegetated land cover.



```
1 X
✓ 0.0s  Open 'X' in Data Wrangler
```

	B02	B03	B04	B08	B8A
119500	1429.0	1659.0	1682.0	3012.0	3186.0
10574	312.0	569.0	275.0	3864.0	4215.0
91995	1394.0	1566.0	1524.0	3060.0	3169.0
112423	1377.0	1624.0	1575.0	3454.0	4184.0
22479	176.0	431.0	196.0	3414.0	4098.0
...	...	...	...	...	...
95588	1196.0	1466.0	1230.0	4044.0	3988.0
19227	265.0	468.0	229.0	3890.0	4787.0
91553	1207.0	1386.0	1237.0	3662.0	3742.0
127515	1280.0	1529.0	1284.0	4124.0	4541.0
99978	1536.0	1662.0	1404.0	1660.0	2744.0

```
1 y
✓ 0.0s  Open 'y' in Data Wrangler
119500    4
10574     1
91995     4
112423    2
22479     1
          ..
95588     1
19227     1
91553     1
127515    2
99978     3
Name: label, Length: 4412, dtype: int64
```

Figure 3 *Matrix of datasets*

In the initial run, a supervised classification using Random Forest (RF) processes seven land-cover classes (0: no data/unclassified; 1: dark area; 2: high vegetation; 3: shrub; 4: water body; 5: cloud; and 6: cannabis plantation). This RF-based classification achieved an overall accuracy of 96.89% and a Kappa coefficient of 0.7768, which falls within the $0.6 < k < 0.8$ range, indicating substantial agreement. However, the resulting confusion matrix, as shown in Figure 4(a), indicates that the classification did not yield optimal results for land cover mapping. The high classification accuracy is primarily due to the model's tendency to predict high vegetation cover (class 2), with 28,383 pixels correctly classified as True Positives (TP). The model failed to detect all cannabis plantations, as illustrated in Figure 4(b) where manually digitized cannabis areas (light blue) are compared with the predicted cannabis pixels (white). The model successfully identified only two out of three cannabis plantations, with non-cannabis pixels misclassified as cannabis. In particular, 185 cannabis pixels were incorrectly identified as high vegetation and 12 pixels as clouds, whereas just 27 pixels were accurately identified as cannabis (TP).These results indicate the model's poor performance in detecting cannabis plantations, likely due to unbalanced training samples causing overfitting.

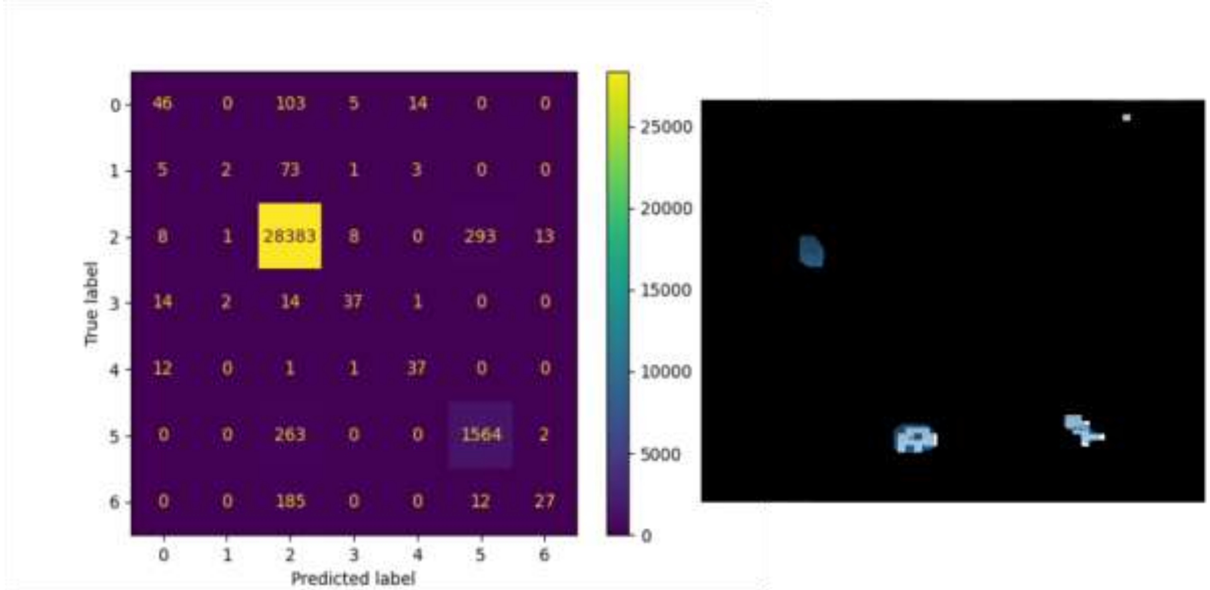


Figure 4. *RF-based 7-Classes Classification Result: (a) Confusion Matrix and (b) Map of detected Cannabis Plantation*

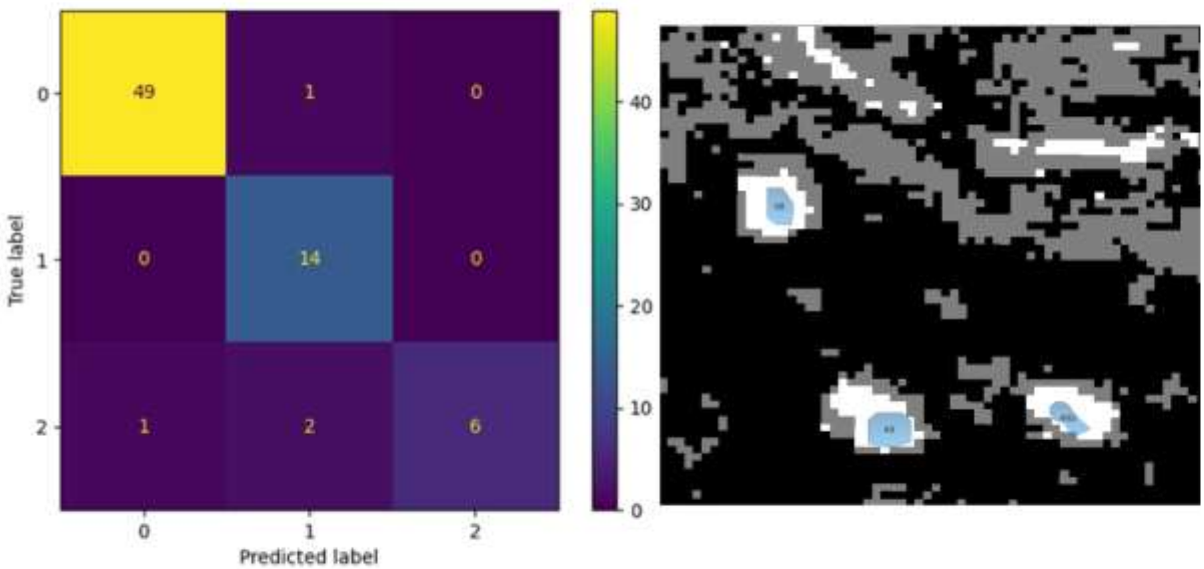


Figure 5. *RF-based 3-Classes Classification Result: : (a) Confusion Matrix and (b) Map of detected Cannabis Plantation*

By refining dataset pixels selection and focusing on three dominant land-cover classes (0: high vegetation/forest, 1: shrub, and 2: cannabis plantation), a supervised classification was re-conducted using only these classes. The confusion matrix from the RF-based classification on test data (Figure 5(a)) indicates a strong performance, with an overall accuracy of 94.52% and a Kappa coefficient of 0.9283. Spatially, as shown in Figure 5(b), all cannabis plantation areas (illustrated in white colour) were effectively detected, though some shrub pixels were misclassified as cannabis, appearing as scattered white pixels among grey pixels. Additional accuracy metrics further support the model's performance, with an average precision of 91.14%, an average recall of 95.25%, and an F1 Score of 0.9315.

To enhance cannabis plantation detection and reduce misclassification of shrubs, hyperparameter tuning was performed. The model retained three classes—forest (black), shrub (grey), and cannabis plantation (white)—and generated the confusion matrix in Figure 6(a). Forest, shrub, and cannabis were labelled as 0, 1, and 2, respectively. The model achieved 90.18% overall accuracy, with a Kappa coefficient of 0.8309. Additional metrics included an average precision of 83.23%, recall of 94.62%, and an F1 score of 0.8856; it showed that the performance of this model is acceptable. It is noted that for this model, the values of overall accuracy, Kappa coefficient, and other performance metrics declined. Spatially, this model outperformed previous model by reducing shrub misclassification as cannabis plantations. The spatial classification results (Figure 6b) depict land cover distribution: forest (black), shrub (grey), and cannabis (white). The model accurately identified cannabis plantations across three test locations with reduced spatial misclassification. Compared to Figure 5(b), shrub misclassification as cannabis was significantly reduced.

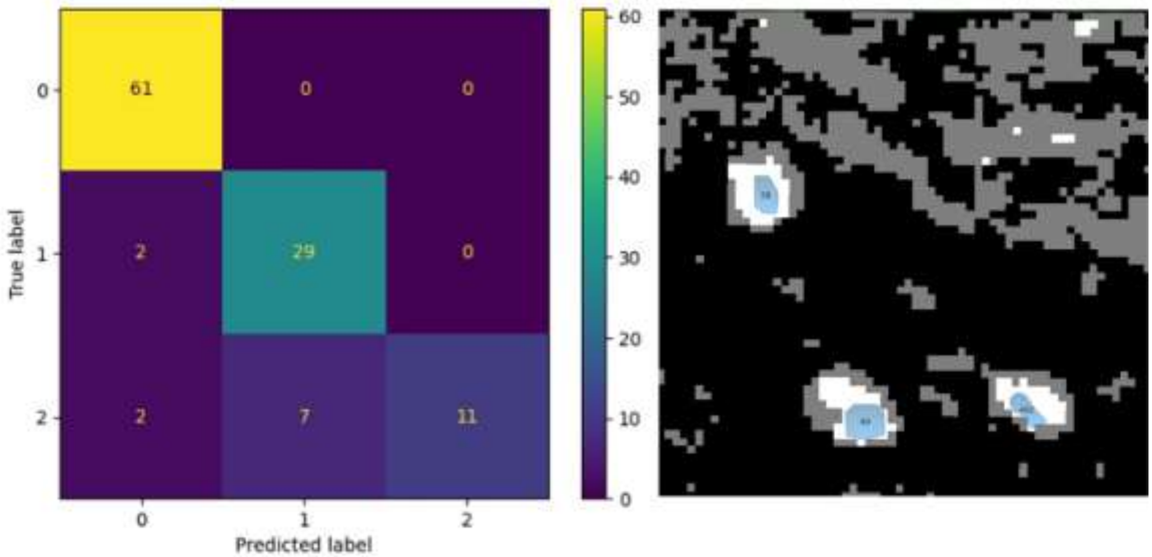


Figure 6 *Improved RF-based 3-Classes Classification Result: (a) Confusion Matrix and (b) Map of detected Cannabis Plantation*

Using the GBT algorithm for supervised classification, the model categorizes test data into four land-cover classes. The confusion matrix (Figure 7a) indicates 84.89% overall accuracy and a Kappa value of 0.7712, with an average precision of 84.88%, recall of 85.57%, and an F1-score of 0.8522. However, cannabis plantations are frequently misclassified as high vegetation or shrubs. Figure 7(b) illustrates the spatial distribution of land-cover predictions: forest (violet), shrub (teal), water body (red), and cannabis plantation (chartreuse). Despite detecting cannabis plantations (label 69), shrub misclassification remains prevalent. Hyperparameter tuning did not improve model performance. Compared to the RF classifier, the GBT model shows also lower values of overall accuracy, Kappa, precision, recall, and F1-score.

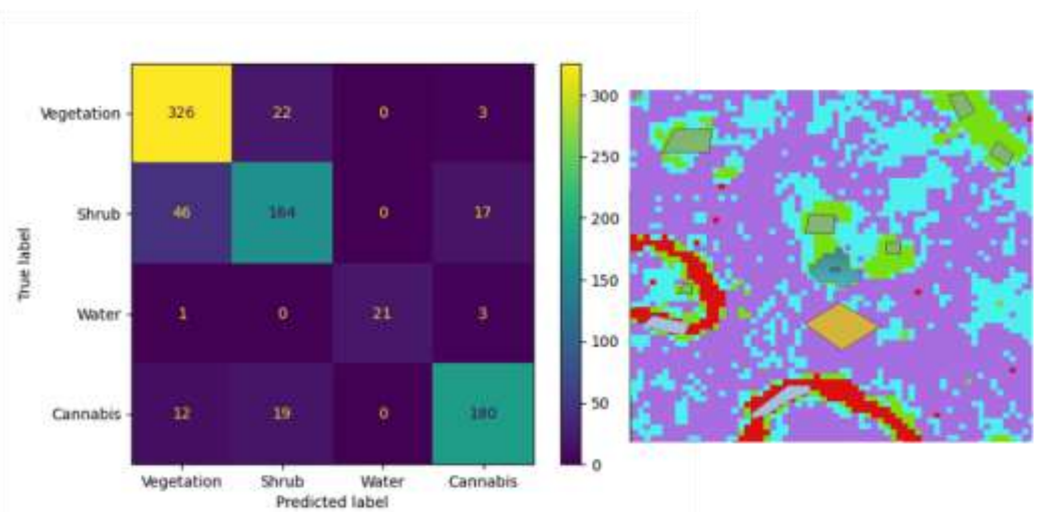


Figure 7 GBT-based 4-Classes Classification Result: (a) Confusion Matrix and (b) Map of detected Cannabis Plantation

The models developed in this study show achieved an overall accuracy exceeding 80% meeting criteria for a reliable model[1]. Additionally, these models also yield Kappa values above 0.7 indicate substantial agreement (Viera & Garret, 2005). High cloud cover in the study area hindered dataset development, necessitating future study on development of data fussion for multi-mission, multi-sensor remote sensing data and statistical features to mitigate data limitations. Moreover, as these models are based on monoculture cannabis plantations, studies incorporating hyperspectral satellite data for spectral signature detection should be explored. Further investigation into alternative machine learning and deep learning algorithms may also enhance model performance

Conclusion

The proposed model achieves 90.18% overall accuracy with a Kappa coefficient of 0.8309 for the improved RF-based model and 84.89% accuracy with a Kappa coefficient of 0.7712 for the GBT-based model. The RF model outperforms the GBT model in both accuracy and spatial detection. With high accuracy and precise location of cannabis plantations, both models offer a faster and more effective approach for detecting illegal plantations compared to traditional law enforcement methods.

Development of cannabis plantation detection was limited by the lack of cloud-free data. Future research should utilize multi-mission and multi-resolution datasets to improve data availability. The model's inability to detect polyculture plantations should also be addressed for future research. Hyperspectral satellite data could enhance polyculture detection by leveraging the unique spectral signature of cannabis plants. Additionally, exploring other machine learning and deep learning algorithms may further optimize detection performance.

4. Acknowledgement

This manuscript would not have been possible without the contributions of my colleagues at the Research Center for Geoinformatics, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN). I sincerely thank Fauzan Al-Ayubi, Olivia Maftukhaturrizqoh, Dony Kushardono, Syarif Budhiman, Karunika Diwyacitta, Bayu S. Adhitama, and Ika S. Supriyani for their support in software development, data preparation, and valuable discussions.

5. References

Armenteras, D., Rodríguez, N., Retana, J., & Morales, M. (2011). Understanding deforestation in montane and lowland forests of the Colombian Andes. Regional Environmental Change, 11(3), 693–705. https://doi.org/10.1007/s10113-010-0200-y

Bicakli, F., Kaplan, G., & Alqasemi, A. S. (2022). Cannabis sativa L. Spectral Discrimination and Classification Using Satellite Imagery and Machine Learning. Agriculture (Switzerland), 12(6). https://doi.org/10.3390/agriculture12060842

Breiman, L. (2001). Random Forests (Vol. 45).

Corea, F. (2017). Artificial Intelligence and Exponential Technologies: Business Models Evolution and New Investment Opportunities (J. Kacprzyk, Ed.). SpringerBriefs in Applied Sciences and Technology. http://www.springer.com/series/10618

Dillis, C., McIntee, C., Butsic, V., Le, L., Grady, K., & Grantham, T. (2020). Water storage and irrigation practices for cannabis drive seasonal patterns of water extraction and use in Northern California. Journal of Environmental Management, 272. https://doi.org/10.1016/j.jenvman.2020.110955

Dimov, D., Low, F., Ibrakhimov, M., Stulina, G., & Conrad, C. (2017). SAR and optical time series for crop classification. IEEE Xplore. https://doi.org/10.1109/IGARSS.2017.8127076

Fang, P., Zhang, X., Wei, P., Wang, Y., Zhang, H., Liu, F., & Zhao, J. (2020). The classification performance and mechanism of machine learning algorithms in winter wheat mapping using Sentinel-2 10 m resolution imagery. Applied Sciences (Switzerland), 10(15).

[1] https://www.deepchecks.com/question/what-is-a-good-accuracy-score-in-machine-learning/

https://doi.org/10.3390/app10155075
Ferreira, A., Felipussi, S. C., Pires, R., Avila, S., Santos, G., Lambert, J., Huang, J., & Rocha, A. (2019). Eyes in the skies: A data-driven fusion approach to identifying drug crops from remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(12), 4773–4786. https://doi.org/10.1109/JSTARS.2019.2917024
Viera, A. J., & Garret, J. M. (2005). Interrater_agreement.Kappa_statistic. Family Medicine, 37(5), 360–363.
Khan, W., Minallah, N., Khan, I. U., Wadud, Z., Zeeshan, M., Yousaf, S., & Qazi, A. B. (2020). On the Performance of Temporal Stacking and Vegetation Indices for Detection and Estimation of Tobacco Crop. IEEE Access, 8, 103020–103033. https://doi.org/10.1109/ACCESS.2020.2998079
Mattiuzzi, M., Bussink, C., & Bauer, T. (2014). Analysing Phenological Characteristics Extracted from Landsat NDVI Time Series to Identify Suitable Image Acquisition Dates for Cannabis Mapping in Afghanistan. https://doi.org/10.1127/1432-8364/2014/02311
Morello, V., Brousseau, V. D., Wu, N., Wu, B. Sen, MacPherson, S., & Lefsrud, M. (2022). Light Quality Impacts Vertical Growth Rate, Phytochemical Yield and Cannabinoid Production Efficiency in Cannabis sativa. Plants, 11(21). https://doi.org/10.3390/plants11212982
PolyGeo. (2022, February 14). Mask Sentinel-2 image using SCL product in Google Earth Engine. Https://Gis.Stackexchange.Com/Questions/423823/Mask-Sentinel-2-Image-Using-Scl-Product-in-Google-Earth-Engine. https://gis.stackexchange.com/questions/423823/mask-sentinel-2-image-using-scl-product-in-google-earth-engine
Putri, W. U., Pinuri, W., Bastara, Prastya, A., Indriany, H. S., & Wulan, L. W. (2022). Indonesia Drugs Report 2022. In PUSDATIN-BNN.
Rodriguez-Galiano, V. F., & Chica-Rivas, M. (2014). Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models. International Journal of Digital Earth, 7(6), 492–509. https://doi.org/10.1080/17538947.2012.748848
Sirro, L., Häme, T., Rauste, Y., Kilpi, J., Hämäläinen, J., Gunia, K., de Jong, B., & Pellat, F. P. (2018). Potential of different optical and SAR data in forest and land cover classification to support REDD+ MRV. Remote Sensing, 10(6). https://doi.org/10.3390/rs10060942
Sujud, L., Jaafar, H., Haj Hassan, M. A., & Zurayk, R. (2021). Cannabis detection from optical and RADAR data fusion: A comparative analysis of the SMILE machine learning algorithms in Google Earth Engine. Remote Sensing Applications: Society and Environment, 24. https://doi.org/10.1016/j.rsase.2021.100639
What Is Random Forest? | IBM. (n.d.). Retrieved February 12, 2025, from https://www.ibm.com/think/topics/random-forest
Zhang, H., Zhang, Y., Liu, K., Lan, S., Gao, T., & Li, M. (2023). Winter wheat yield prediction using integrated Landsat 8 and Sentinel-2 vegetation index time-series data and machine learning algorithms. Computers and Electronics in Agriculture, 213, 108250. https://doi.org/10.1016/j.compag.2023.108250
Zhu, L., Sun, W., Zhang, Q., Wang, C., & Guo, Z. (2023). Fine-grained agricultural and pastoral information extraction using Sentinel-1 and Sentinel-2 intra-year time series in Jingyang District, Deyang City. Advances in Space Research, 72(9), 4031–4047. https://doi.org/10.1016/j.asr.2023.07.061