Comparative Analysis of Machine Learning Models for Sentinel-2 Based Classification of Bornean Heath Forest

1st Dwi Ahmad Dzulhijjah
Magister of Informatics, Universitas Amikom Yogyakarta, Yogyakarta Indonesia | Department of Informatics and Computer Engineering, Politeknik Elektronika Negeri Surabaya, Sukolilo, Surabaya, Indonesia
dwiahmadz@students.amikom.ac.id

2nd Kusrini
Informatics, Postgraduate Program, Universitas AMIKOM Yogyakarta, Yogyakarta, Indonesia | Fundación para la Investigación y Desarrollo Tecnológico de la Sociedad del Conocimiento, Murcia, Spain
kusrini@amikom.ac.id (corresponding author)

3rd Rodrigo Martínez-Béjar
Informatics, Faculty of Computer Science, University of Murcia, Murcia, Spain | Fundación para la Investigación y Desarrollo Tecnológico de la Sociedad del Conocimiento, Murcia, Spain
rodrigo@um.es

Received: 1 January 202X | Revised: 2 February 202X | Accepted: 3 March 202X ("ETASR dates" style)

 DOI: https://doi.org/10.48084/etasr.XXXX

Abstract

Bornean heath forests, known as *hutan kerangas*, are fragile ecosystems facing significant anthropogenic threats. This study utilizes Sentinel-2 satellite imagery combined with machine learning models to achieve accurate classification of these forests. The Random Forest and Gradient Tree Boost models demonstrated the highest performance, attaining overall accuracy scores of 96.66% and 96.69%, respectively, with Kappa coefficients of 0.945. A 5-fold cross-validation process further confirmed the robustness and reliability of the models. The findings highlight the significant potential of integrating remote sensing technologies with advanced machine learning techniques for scalable ecosystem monitoring. This approach supports conservation efforts and promotes sustainable management of Bornean heath forests in response to escalating environmental pressures.

Keywords-Heath Forest Classification; Sentinel-2 Imagery; Machine Learning Models; Remote Sensing; Climate Action

I. Introduction

Bornean heath forests, locally known as *hutan kerangas* [1], [2], are unique and fragile ecosystems that thrive in nutrient poor, acidic, and quartz-sand-dominated soils [3]. Despite their ecological resilience, these forests face severe threats from anthropogenic activities such as uncontrolled forest fires, sand mining, illegal logging, and land use conversion [4]. Between 2009 and 2024, approximately 32,895,682.96 m2 of heath forests have been deforested, reflecting an alarming rate of habitat loss. This deforestation contributes to global carbon emissions, biodiversity decline, and hydrological imbalances, exacerbating local and global climate challenges.

Heath forest ecosystems play a vital ecological role, including carbon sequestration, maintaining endemic biodiversity, and supporting plant species such as *Nepenthes gracilis* and *Rhodomyrtus tomentosa* [5], [6]. These forests are also critical for regulating hydrological cycles and providing economic resources for local communities. However, their fragmented distribution and vulnerability to degradation have made monitoring and management increasingly difficult. Traditional methods, such as manual surveys and unsupervised classification with remote sensing, are often inadequate to capture the heterogeneity and dynamic changes within these ecosystems [7].

In recent years, remote sensing technologies have proven effective in monitoring forest ecosystems and land cover changes. Platforms like Landsat and Sentinel-2, when integrated with Geographic Information Systems (GIS), offer scalable solutions for mapping and monitoring forest areas [8], [9]. For instance, a previous study successfully mapped heath forest areas in East Kalimantan, Indonesia, using Landsat-9 data and unsupervised classification techniques [1]. However, limitations remain in spatial resolution, spectral details, and classification accuracy, particularly in heterogeneous and fragmented ecosystems like heath forests.

To overcome these challenges, machine learning (ML) models have emerged as robust alternatives for land cover classification. Techniques such as Random Forest (RF), Support Vector Machines (SVM), Gradient Tree Boost (GTB), and K-Nearest Neighbors (KNN) have shown superior performance in handling high-dimensional remote sensing data [10], [11]. These models, when applied to high-resolution Sentinel-2 imagery, can effectively identify subtle variations in vegetation structure, canopy density, and soil characteristics. Sentinel-2, with its improved spatial (10-20 m) and spectral resolution, provides a significant advantage over Landsat data for detailed ecosystem classification [12].

This study aims to advance heath forest classification methodologies by leveraging Sentinel-2 satellite data and evaluating the performance of multiple machine learning models. Specifically, the objectives of this research are as follows:

- This study aims to evaluate and compare the classification performance of machine learning models, focusing on their ability to capture the spatial and spectral complexity of heath forest ecosystems using Sentinel-2 data.
- To demonstrate the effectiveness of Sentinel-2 imagery for capturing finer spatial and spectral details of heath forest ecosystems compared to conventional approaches.
- To provide a robust and replicable methodology for monitoring heath forest changes using machine learning and high-resolution remote sensing data.

The contributions of this study are significant for the monitoring and conservation of heath forests. By integrating Sentinel-2 imagery with advanced machine learning techniques, this research enhances the precision and reliability of forest classification, achieving high classification accuracies exceeding 94% across models. The comparative evaluation of multiple models offers valuable insights into their strengths and limitations, identifying the most suitable approach for classifying heterogeneous and fragmented ecosystems. These findings provide a foundation for developing scalable and data driven monitoring systems, supporting conservation efforts, policymaking, and sustainable management of Bornean heath forests amidst increasing environmental pressures.

II. Related Works

The accurate classification and mapping of land use and land cover (LULC) have been critical areas of research [13], particularly in regions with complex and heterogeneous ecosystems such as heath forests. These studies often integrate advanced remote sensing technologies, machine learning algorithms, and ecological insights to enhance classification accuracy and utility [14].

Multi-temporal remote sensing has emerged as a powerful approach for LULC classification, enabling the detection of seasonal and phenological changes that are otherwise indistinguishable in single-date imagery [15]. A study utilizing Sentinel-2 multi-temporal datasets combined with spectral indices such as NDVI and MNDWI demonstrated significant advantages for LULC classification in areas with complex ecosystems [16]. By leveraging Random Forest (RF) and Support Vector Machine (SVM) models, the study achieved over 90% accuracy, emphasizing the importance of incorporating temporal dynamics in modeling efforts. Similarly, research on cropland mapping highlighted that multi-temporal approaches, when paired with ensemble methods like RF, could significantly improve classification accuracy by addressing seasonal variations in vegetation cover [17].

In Indonesia, research efforts have focused on understanding the characteristics and spatial distribution of heath forests (Kerangas), particularly in the context of the country's new capital city, Nusantara. A study employed Landsat 9 imagery alongside GIS-based methods to map heath forests, using parameters such as elevation, soil pH, and NDVI [1]. This study, while utilizing traditional unsupervised classification techniques, provided foundational insights into the ecological and spatial dynamics of Kerangas forests. These findings align with earlier works in Southeast Asia [18], which emphasized the role of nutrient-poor soils and unique vegetation compositions in defining these ecosystems.

Sentinel-2 provides significant advantages over platforms like Landsat, including higher spatial resolution (10–20 m) and an expanded spectral range [19]. These characteristics enable the detection of subtle vegetation and soil variations essential for distinguishing heath forest types. The improved temporal resolution also allows for more frequent monitoring, critical for dynamic ecosystems like heath forests. This capability underscores Sentinel-2's unique contribution to advancing ecosystem classification efforts.

Building on these advancements, this study aims to refine the classification of Bornean heath forests using a combination of traditional and advanced machine learning methods. By integrating temporal and spectral data, this research seeks to address the unique challenges posed by the sparse vegetation, nutrient-poor soils, and diverse ecological conditions of Kerangas ecosystems.

III. METHODOLOGY

A. Research Framework

This study adopts a structured framework for the classification of Bornean heath forests using Sentinel-2 satellite imagery and machine learning techniques. The framework consists of five stages: on-site observation, data collection and annotation, data preprocessing, model training and validation, and performance evaluation. This iterative process integrates remote sensing data with machine learning models for classifying heath forest regions. The workflow ensures accurate data representation and model performance validation. Figure 1 illustrates the overall methodology.

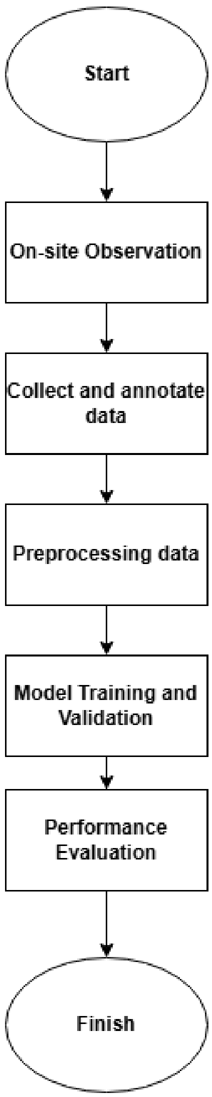

Fig. 1. Overall Workflow of the Research Framework for Heath Forest Classification.

B. On-site Observation

On-site observations were conducted to validate the alignment between satellite imagery and ground conditions across 30 points in Kubu, Kumai, and Kotawaringin Barat Regency. Field visits, conducted over 12 sessions between May and November, confirmed three classes of heath forest: Kerangas Kolam (173 samples), Kerangas Hutan (14 samples), and Kerangas Rawa (9 samples), resulting in a total of 209 annotated in-land samples. These observations ensured the dataset's accuracy by confirming the suitability of Sentinel-2 satellite imagery for representing ecological features.

The documentation from the field observations provided essential evidence for the distinct characteristics of each class, which forms the foundation for model training and validation. Figure 2 illustrates examples of field observations, highlighting key ecological traits corresponding to the annotated classes.

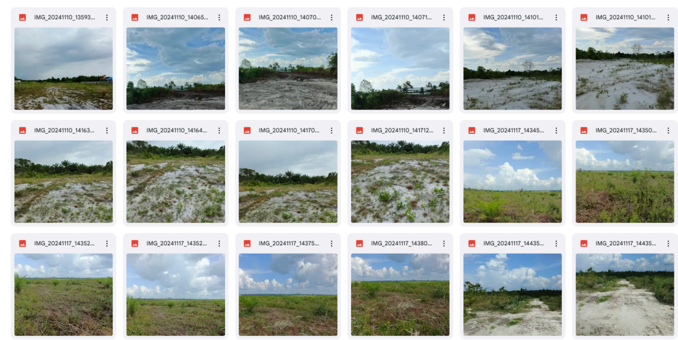

Fig. 2. Field observation documentation for heath forest classification.

C. Data Annotation and Preprocessing

The first step involves the annotation of training regions for heath forests using representative samples, namely Kerangas Kolam, Kerangas Rawa, and Kerangas Hutan. These regions are selected and labeled with class properties to represent distinct heath forest types. Sentinel-2 imagery, covering the period from January 2024 to November 2024, is used as the primary input data. Table I details the key preprocessing steps applied to the Sentinel-2 imagery.

TABLE I. DATA PREPROCESSING STEPS FOR SENTINEL-2 IMAGERY

Step	Table Column Head
Cloud Filtering	Exclude images with CLOUDY_PIXEL_PERCENTAGE > 10%.
Image Scaling	Scale pixel values by a factor of 0.0001 to ensure compatibility with the reflectance range.
Compositing and Clipping	A median composite of filtered images is generated to reduce noise, and the resulting image is clipped to the Region of Interest (ROI).
Visualization	The preprocessed image is visualized using the True Color Composite bands (B4, B3, B2) to assess its quality before analysis.

After completing the data preprocessing steps outlined in Table I, Regions of Interest (ROI) were delineated for each class of heath forests: Kerangas Kolam, Kerangas Rawa, and Kerangas Hutan. These ROIs represent diverse ecological characteristics and serve as the foundation for annotated training data.

Figure 3 showcases a visual representation of the preprocessed Sentinel-2 imagery, highlighting the ROIs for each class. The True Color Composite bands (B4, B3, B2) were used to visualize the distinct features of these regions, ensuring the clarity and precision of the annotated samples.

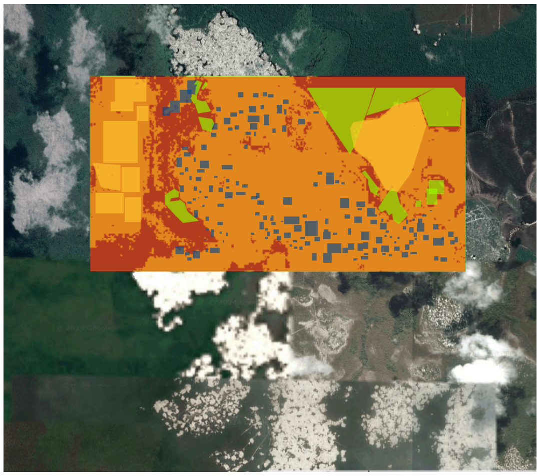

Fig. 3. Regions of Interest (ROI) for heath forest classes.

D. Feature Selection and Data Preparation

The spectral bands used for classification include Sentinel-2 bands: B2, B3, B4, B5, B6, B7, B8, B9, B11, and B12. These bands serve as input features for machine learning models. Table II provides detailed information on the selected spectral bands, including their wavelengths and spatial resolutions.

TABLE II. SELECTED SENTINEL-2 SPECTRAL BANDS FOR FEATURE INPUT

Band	Wavelength (nm)	Spatial Resolution (m)
B2 (Blue)	490	10
B3 (Green)	560	10
B4 (Red)	665	10
B5 (Red Edge 1)	705	20
B6 (Red Edge 2)	740	20
B7 (Red Edge 3)	783	20
B8 (NIR)	842	10
B9 (Water Vapor)	945	60
B11 (SWIR 1)	1610	20
B12 (SWIR 2)	2190	20

The labeled training data is generated by sampling the preprocessed image over annotated regions with the following properties:

- **Class Labels**: Defined classes for heath forest regions.
- **Spatial Resolution**: 10-meter scale to match Sentinel-2 spatial resolution.

The dataset is randomly split into two subsets: 80% for training and 20% for testing, ensuring balanced representation across classes.

E. Machine Learning Model Training and Classification

The core of this study involves the implementation and comparison of multiple machine learning models for heath forest classification. Table III summarizes the machine learning models used, their key parameters, and respective advantages.

TABLE III. MACHINE LEARNING MODELS USED FOR CLASSIFICATION

Model	Key Parameters and Advantages
Random Forest (RF)	100 decision trees; Handles high-dimensional data effectively.
Support Vector Machine (SVM)	Linear kernel; Suitable for small datasets.
Gradient Tree Boost (GTB)	50 iterations; Robust against overfitting.
Classification and Regression Trees (CART)	Depth: 5; Trees: 100; Produces interpretable results.
K-Nearest Neighbors (KNN)	5 neighbors; Euclidean distance; Simple and non- parametric.

For each model, the following steps are executed:

1. **Training**: Models are trained using the 80% training data and the selected spectral bands as input features.
2. **Classification**: The trained models classify the Sentinel-2 image into heath forest classes.
3. **Validation**: The remaining 20% testing data is used to validate the classification results, generating confusion matrices and overall accuracy metrics.

F. K-Fold Cross Validation

To further evaluate the robustness of each model, a 5-Fold Cross Validation (K-Fold CV) is performed. The process involves the following steps:

- The labeled dataset is randomly partitioned into 5 equal folds.
- For each fold, one subset serves as the testing data while the remaining four subsets are used for training.
- Models are trained and evaluated iteratively over all five folds.
- The accuracy and Kappa statistics are averaged across the folds to assess overall performance.

The use of K-Fold CV ensures that the evaluation is unbiased and accounts for variability in the data distribution.

G. K-Fold Cross Validation

The performance of each model is quantified using the following metrics:

- **Confusion Matrix**: Provides a detailed breakdown of true and predicted classifications.

- **Overall Accuracy**: Measures the proportion of correctly classified samples.
- **Kappa Coefficient**: Evaluates the agreement between predicted and actual classifications while accounting for chance.

The results are compared across models to determine the most effective technique for heath forest classification.

H. Experimental Workflow Summary

The experimental workflow can be summarized as follows:

1) Sentinel-2 data collection, preprocessing, and annotation of training regions.
2) Feature extraction using Sentinel-2 bands and data splitting into training and testing sets.
3) Implementation of machine learning models (RF, SVM, GTB, CART, and KNN).
4) Validation using confusion matrices, overall accuracy, and K-Fold Cross Validation.
5) Comparative analysis of model performance and identification of the best-performing classifier.

The outcomes of this methodology provide a robust and scalable approach for the classification and monitoring of heath forests using Sentinel-2 satellite imagery.

IV. Results and Discussion

A. Understanding Model Selection

The selection of models such as Random Forest (RF) and Gradient Tree Boost (GTB) was driven by their ability to handle high dimensional spectral data effectively, which is critical when working with Sentinel-2 imagery. Unlike linear models such as Support Vector Machines (SVM), ensemble methods like RF and GTB leverage multiple decision trees to enhance classification precision, making them well suited for the complexity of heath forest ecosystems.

B. Classification Accuracy

The performance of the machine learning models for the classification of Bornean heath forests is summarized in Table IV. All models demonstrated high classification accuracies, with the Random Forest (RF) and Gradient Tree Boost (GTB) models achieving the highest accuracy scores of over 96%.

TABLE IV. PERFORMANCE METRICS FOR HEATH FOREST CLASSIFICATION

Model	Overall Accuracy (%)	Kappa Coefficient
Random Forest (RF)	96,66	0,945
Support Vector Machine (SVM)	94,09	0,9
Gradient Tree Boost (GTB)	96,69	0,945
Classification and Regression Trees (CART)	95,98	0,932
K-Nearest Neighbors (KNN)	96,13	0,938

The Random Forest and Gradient Tree Boost models consistently outperformed other classifiers, showcasing their ability to effectively handle the high dimensional spectral data of Sentinel-2 imagery. The Kappa coefficients, exceeding 0.93 for all models, indicate strong agreement between predicted and true labels, emphasizing the reliability of the classifications.

C. K-Fold Cross Validation

The performance of the machine learning models for the classification of Bornean heath forests is summarized in Table IV. All models demonstrated high classification accuracies, with the Random Forest (RF) and Gradient Tree Boost (GTB) models achieving the highest accuracy scores of over 96%.

TABLE V. PERFORMANCE METRICS FOR HEATH FOREST CLASSIFICATION

Model	Overall Accuracy (%)	Kappa Coefficient
Random Forest (RF)	96.48	0,945
Support Vector Machine (SVM)	93.56	0,9
Gradient Tree Boost (GTB)	96.43	0,944
Classification and Regression Trees (CART)	95.59	0,932
K-Nearest Neighbors (KNN)	95.98	0,938

The results confirm the robustness of the Random Forest and Gradient Tree Boost models, with consistent performance across different data folds. The slight variations in accuracy and Kappa scores across folds highlight the natural variability in the dataset, but the overall high metrics demonstrate the reliability of the approach.

D. Discussion

The study highlights the potential of combining Sentinel-2 imagery with machine learning models for accurate classification of heath forest ecosystems. Among the evaluated models, the Random Forest and Gradient Tree Boost classifiers showed superior performance. These ensemble methods leverage multiple decision trees to enhance classification precision, making them particularly effective for high-dimensional spectral data such as Sentinel-2 imagery.

The K-Nearest Neighbors (KNN) model also demonstrated commendable performance despite its simplicity. This suggests that the Sentinel-2 spectral features are highly discriminative, making KNN a viable option for applications where computational simplicity is a priority.

The Support Vector Machine (SVM) model exhibited slightly lower accuracy, which could be attributed to its reliance on linear decision boundaries. However, SVM remains a robust choice for smaller datasets or preliminary analyses due to its computational efficiency and strong theoretical foundations.

E. Implications for Conservation and Land Use Policy

The high classification accuracies achieved in this study underscore the potential for developing operational systems to monitor heath forests. Such systems could support conservation strategies by identifying areas at risk of degradation, guide land-use planning by providing detailed ecosystem maps, and inform policymakers in mitigating climate change impacts through evidence-based decisions. These applications enhance the practical relevance of the study's findings.

F. Implications for Heath Forest Monitoring

The findings of this study underscore the advantages of using high-resolution Sentinel-2 imagery for monitoring Bornean heath forests. The improved spatial and spectral resolution of Sentinel-2 enables the detection of subtle variations in vegetation and soil characteristics, which are critical for distinguishing between different heath forest types. Integrating this data with machine learning models provides a scalable, data-driven methodology for monitoring fragmented and heterogeneous ecosystems.

The high classification accuracies achieved in this study provide a strong foundation for developing operational systems to monitor heath forests. Such systems can inform conservation strategies, guide land-use planning, and support policymakers in mitigating the impacts of anthropogenic activities on these fragile ecosystems.

G. Limitations and Future Work

This study achieved promising results; however, several limitations warrant consideration. First, classification accuracy may be influenced by changes in data quality, such as increased cloud cover or seasonal variability in Sentinel-2 imagery. Second, the study focused on three primary heath forest types, which may not fully capture the diversity of heath forest ecosystems.

Future research should address these limitations by incorporating multitemporal Sentinel-2 data to better account for seasonal dynamics and variability in vegetation characteristics. Expanding the geographic scope of training data and including additional heath forest types could enhance model generalizability. Ancillary datasets, such as topographic features, climatic variables, and soil properties, could also be integrated to provide additional context for classification.

Moreover, exploring advanced deep learning approaches, such as Convolutional Neural Networks (CNNs), offers a compelling avenue for improving classification performance. CNNs excel in extracting spatial and spectral features and could leverage the rich Sentinel-2 data to identify complex patterns in fragmented ecosystems. Future studies should also evaluate the performance of CNN-based methods in comparison to traditional machine learning models to determine their relative strengths and limitations.

Finally, to enhance the robustness of classification systems, integrating data fusion techniques with complementary remote sensing platforms and testing models across diverse environmental conditions will be critical for operational scalability and wider applicability.

H. Conclusion

This study demonstrates the effectiveness of integrating Sentinel-2 imagery with machine learning models for classifying heath forest ecosystems. The results highlight the potential of these approaches to support large-scale monitoring and conservation efforts, ensuring the sustainability of these ecologically significant forests amidst increasing environmental pressures.

V. Replication Package

To ensure reproducibility, all scripts, datasets, and model configurations used in this study are provided in the replication package available at [20]. This package includes the Sentinel-2 preprocessing steps, annotated training regions, model training scripts, and evaluation metrics. Users can follow the provided guidelines to replicate the experiments or adapt the methodology for other land cover classification tasks.

VI. Threats to Validity

While the findings are promising, several threats to validity exist. The reliance on Sentinel-2 imagery may introduce biases due to seasonal variations or cloud cover, which were mitigated through strict preprocessing steps but could still influence the results. Additionally, the models were trained on specific heath forest types, limiting their generalizability to other ecosystems or regions.

Another potential threat is the variability in training data quality. Although K-Fold Cross Validation was employed to ensure robustness, the small sample size for certain classes could affect model reliability. Future work should address these limitations by incorporating larger and more diverse datasets, as well as evaluating the models under different environmental conditions.

VII. Conclusion

This study demonstrates the effectiveness of combining Sentinel-2 imagery with machine learning models for the classification of Bornean heath forest ecosystems. The results highlight the robust performance of Random Forest and Gradient Tree Boost classifiers, both achieving overall accuracies above 96% and strong Kappa coefficients, indicating high reliability. The application of 5-Fold Cross Validation further confirmed the robustness of these models across varied data partitions.

The study underscores the potential of using high-resolution remote sensing data to detect subtle ecological variations, enabling scalable and precise monitoring of fragmented ecosystems. These findings are instrumental for conservation planning, land-use management, and policy making aimed at preserving heath forests amidst escalating anthropogenic pressures. Future work should explore temporal dynamics, additional forest classes, and advanced deep learning models to further enhance classification accuracy and generalizability

Acknowledgment

his research was funded by the IEEE Student Branch of Universitas AMIKOM Yogyakarta under the IEEE ASB Research Grant program for the 2024 fiscal year,

References

[1] R. M. Nisaa, U. K. Sari, and Y. B. Sulistioadi, "Heath forest identification using remote sensing in the surrounding area of the new capital city of Indonesia," in *IOP Conference Series: Earth and Environmental Science*, vol. 1266, no. 1, p. 012068, 2023, doi: 10.1088/1755- 1315/1266/1/012068.

[2] M. A. Soendjoto, M. K. Riefani, and S. Diana, "Floristic diversity and composition of Kuala Tambangan Heath Forest in Tanah Laut District, South Kalimantan, Indonesia," *Biodiversitas*, vol. 24, no. 10, pp. 5418–5427, 2023, doi: 10.13057/biodiv/d241024.

[3] I. S. Mohd Ikbal, S. Md Jaafar, N. Ahmad, and R. S. Sukri, "Contrasting environmental drivers of tree community variation within heath forests in Brunei Darussalam, Borneo," *Biodiversity Data Journal*, vol. 12, p.e127919, 2024, doi: 10.3897/BDJ.12.e127919.

[4] D. Oktavia, S. D. Pratiwi, S. Munawaroh, A. Hikmat, and I. Hilwan, "The potential of medicinal plants from heath forest: Local knowledge from Kelubi Village, Belitung Island, Indonesia," *Biodiversitas*, vol. 23, no. 7, pp. 3553–3560, 2022, doi: 10.13057/biodiv/d230731.

[5] M. A. Brovelli, Y. Sun, and V. Yordanov, "Monitoring forest change in the Amazon using multi-temporal remote sensing data and machine learning classification on Google Earth Engine," *ISPRS International Journal of Geo-Information*, vol. 9, no. 10, p. 580, 2020, doi: 10.3390/ijgi9100580.

[6] D. Hoekman et al., "Wide-area near-real-time monitoring of tropical forest degradation and deforestation using Sentinel-1," *Remote Sensing*, vol. 12, no. 19, p. 3263, 2020, doi: 10.3390/rs12193263.

[7] Y. Gao, M. Skutsch, J. Paneque-G´alvez, and A. Ghilardi, "Remote sensing of forest degradation: a review," *Environmental Research Letters*, vol. 15, no. 10, p. 103001, 2020, doi: 10.1088/1748-9326/abaad7.

[8] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, and M. Ranagalage, "Sentinel-2 data for land cover/use mapping: A review," *Remote Sensing*, vol. 12, no. 14, p. 2291, 2020, doi: 10.3390/rs12142291.

[9] K. O. Abduraufovich, "Use of GIS technologies and satellite data for geological research," *Western European Journal of Modern Experiments and Scientific Methods*, vol. 2, no. 10, pp. 24–30, 2024, doi: Not available.

[10] C. Sothe, C. M. De Almeida, M. B. Schimalski, V. Liesenberg, L. E. C. La Rosa, J. D. B. Castro, and R. Q. Feitosa, "A comparison of machine and deep-learning algorithms applied to multisource data for a subtropical forest area classification," *International Journal of Remote Sensing*, vol. 41, no. 5, pp. 1943–1969, 2019, doi: 10.1080/01431161.2019.1681600.

[11] O. Okwuashi and C. E. Ndehedehe, "Deep support vector machine for hyperspectral image classification," *Pattern Recognition*, vol. 103, p. 107298, 2020, doi: 10.1016/j.patcog.2020.107298.

[12] R. H. Topalo˘glu, E. Sertel, and N. Musaoglu, "Assessment of classification accuracies of Sentinel-2 and Landsat-8 data for land cover/use mapping," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-B8, pp. 1055–1059, 2016, doi: 10.5194/isprsarchives-XLI-B8-1055-2016.

[13] J. Wang, M. Bretz, M. A. A. Dewan, and M. A. Delavar, "Machine learning in modelling land-use and land-cover-change (LULCC): Current status, challenges and prospects," *Science of the Total Environment*, vol. 822, p. 153559, 2022, doi: 10.1016/j.scitotenv.2022.153559.

[14] F. Li, T. Yigitcanlar, M. Nepal, K. Nguyen, and F. Dur, "Machine learning and remote sensing integration for leveraging urban sustainability: A review and framework," *Sustainable Cities and Society*, vol. 96, p. 104653, 2023, doi: 10.1016/j.scs.2023.104653.

[15] R. N. Masolele, "Assessing land use following tropical deforestation: Combining remote sensing and deep learning," Ph,D dissertation, Wageningen University, Wageningen, The Netherlands, 2023, doi: 10.18174/583842.

[16] A. Arfa and M. Minaei, "Utilizing multitemporal indices and spec- tral bands of Sentinel-2 to enhance land use and land cover classi- fication with random forest and support vector machine," *Advances in Space Research*, vol. 74, no. 8, pp. 5580–5590, 2024, doi: 10.1016/j.asr.2024.08.062.

[17] A. Tariq, J. Yan, A. S. Gagnon, M. R. Khan, and F. Mumtaz, "Mapping of cropland, cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest," *Geo-spatial Information Science*, vol. 26, no. 3, pp. 302–320, 2023, doi: 10.1080/10095020.2022.2100287.

[18] C. Brown, S. Sj¨ogersten, M. J. Ledger, F. Parish, and D. Boyd, "Remote sensing for restoration change monitoring in tropical peat swamp forests in Malaysia," *Remote Sensing*, vol. 16, no. 15, p. 2690, 2024, doi: 10.3390/rs16152690.

[19] P. S. Nagendram, P. Satyanarayana, and P. R. Teja, "Mapping Paddy Cropland in Guntur District using Machine Learning and Google Earth Engine utilizing Images from Sentinel-1 and Sentinel-2," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12427–12432, Dec. 2023, https://doi.org/10.48084/etasr.6460.

[20] Google Earth Engine Code Repository, "Replica- tion package for land cover classification study," Google Earth Engine, 2024. [Online]. Available: https://code.earthengine.google.com/8fae384fe296584613c96049f3cd15d3. [Accessed: Dec. 19, 2024].

AUTHORS PROFILE

Dwi Ahmad Dzulhijjah is currently a Ph.D. student in Cyber-Physical Systems at Politeknik Elektronika Negeri Surabaya (PENS), Indonesia, focusing on real-time traffic management and anomaly detection. He holds dual master's degrees in Artificial Intelligence from Universitas Gadjah Mada (UGM) and Informatics focusing in Business Intelligence from Universitas AMIKOM Yogyakarta, as well as bachelor's degrees in Biology and Informatics Engineering. Additionally, he has earned a professional engineer degree in Informatics. He is also an AI Engineer at PT. Piranti Kecerdasan Buatan (PIKEBU), specializing in computer vision and edge computing solutions. His research interests include multi-agent systems, distributed AI, and the integration of AI in smart city applications. He can be contacted at dwiahmaddzulhijjah@gmail.com.

Kusrini is a professor from Universitas AMIKOM Yogyakarta Indonesia. She finished her doctoral program in the Computer Science Study Program at Universitas Gadjah Mada Yogyakarta Indonesia in 2010. She is interested in exploring many things related to machine learning and other artificial intelligence fields. She also loves researching decision support systems and databases. She is a member of the IEEE and IEEE Systems, Man, and Cybernetics Society. She can be contacted at email: kusrini@amikom.ac.id. Website: http://kusrini.com/.

Rodrigo Martínez-Béjar is a Professor of Computer Science in the Department of Information and Communication Engineering at the University of Murcia in Spain. He received his Master's in Computer Science from the University of Malaga in 1992 and his Ph.D. from the University of Murcia in Spain in 1997. Also, he holds a Master's degree in Applied Sociology from the University of Murcia. He has held visiting positions at several European, Australian and Latin-American Universities, including the Australian Universities of Adelaide and New South Wales and the French University of Southern Paris. He has served on several Editorial Boards and co-authored more than 100 scientific publications, most of them in peer-reviewed journals and conferences. He has led more than 30 national or international competitive grants and research contracts with industry. In 2006, he became the Head of an institutional, inter-disciplinary University research group. He can be contacted at email: rodrigo@um.es.