

Proceeding Paper

Machine Learning-Based Forest Type Mapping from Multi-Temporal Remote Sensing Data: Performance and Comparative Analysis [†]

Yusuf Ibrahim ^{1,*}, Umar Yusuf Bagaye ² and Abubakar Ibrahim Muhammad ²

¹ Department of Computer Engineering, Ahmadu Bello University, Zaria, Nigeria

² Department of Electrical and Electronics Engineering, Kaduna Polytechnic, Kaduna, Nigeria; uybagaye@gmail.com (U.Y.B.); ibngarko@yahoo.com (A.I.M.)

* Correspondence: yibrahim@abu.edu.ng or yusuf2007ee@gmail.com

[†] Presented at the 5th International Electronic Conference on Remote Sensing, 7–21 November 2023; Available online: <https://ecrs2023.sciforum.net/>.

Abstract: This paper presents a meticulous exploration of advanced machine learning techniques for precise forest type classification using multi-temporal remote sensing data within a woodland environment. The study comprehensively evaluates a diverse range of models, ranging from advanced (ensemble) machine learning (ML) methods to several finely tuned support vector machine (SVM) variants, with a specific focus on Bayesian-optimized svm with radial basis function (RBF) kernel. Our findings highlight the robust performance of the Bayesian-optimized SVM, achieving a high accuracy of up to 94.27%, and average precision and recall of 94.46% and 94.27% respectively. Notably, this accuracy aligns with the levels attained by acclaimed ensemble techniques such as Random Forest and CatBoost while also surpassing those of XGBoost and LightGBM. These results highlight the potential of these methodologies to significantly enhance forest type mapping accuracy compared to traditional (Linear) SVM and black-box neural networks. This, in turn, can enable reliable identification and quantification of key services, including carbon storage and erosion protection, intrinsic to the forest ecosystem. Findings of our comparative study emphasize the profound impact of employing and fine-tuning ML approaches in the realm of remote sensing-based environmental analysis.

Keywords: forest type mapping; remote sensing; machine learning; ensemble learning; support vector machine; bayesian optimization

Citation: Ibrahim, Y.; Bagaye, U.Y.; Muhammad, A.I. Machine

Learning-Based Forest Type

Mapping from Multi-Temporal

Remote Sensing Data: Performance and Comparative Analysis. *Environ. Sci. Proc.* **2023**, *27*, x.

Sci. Proc. **2023**, *27*, x.

<https://doi.org/10.3390/xxxxx>

Academic Editor(s): Name

Published: date



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Accurate identification of different forest types is important for effective environmental management practices. The use of remote sensing technology for classifying the forest types down to individual tree species level has numerous applications, such as sustainable forest management [1], biological studies and surveillance [2], monitoring of invasive species [3] and even advancing the monitoring of sustainable development goals (SDG) [4]. Forests are among the largest reservoir of carbon on land that are most vulnerable to land use and land cover change (LULC) [5] and play a critical role in global carbon cycling and ecological stability. They also contribute in mitigating climate change thus enhancing the environment and ensuring ecological security. Therefore, precise mapping of forest types is crucial for assessing factors like carbon storage, minimizing damage, and enhancing resources. Traditional methods of surveying forests usually involve labor-intensive fieldwork and often fall short when it comes to covering the vast and diverse landscapes of forests. Hence, remote sensing technology emerges as a game-changer and offers a solution to efficient data acquisition in challenging terrains. Specifically, the use of multi-

temporal remote sensing data provides perceptions into the dynamics of forests. Such data collected at different time intervals has the potential to unveil evolving patterns and relationships within the forest ecosystems. Utilizing this potential, along with the capabilities of techniques such as ML offers an innovative approach to accurate forest types mapping. Although previous studies have investigated different remote sensing datasets for this purpose, this research utilized a publicly available dataset comprising data derived from remote sensing investigation that encompass unique spectral attributes within the visible-to-near infrared spectrum carried out using ASTER satellite imagery. The analysis in this research involves a comparative evaluation of multiple machine and ensemble learning algorithms, highlighting their effectiveness in handling multi-temporal data.

2. Related Work

A study by Rodríguez-Galiano et al. (2012) [6] investigated the use of Random Forest (RF) classifier for land cover monitoring through remotely sensed data. The performance of the model on complex land cover classification using Landsat-5 data for 14 categories in Spain yielded 92% accuracy, a Kappa index of 0.92, and good robustness to data reduction and noise. The RF model significantly outperforms a single decision tree in the experiment. Liu et al. (2018) [7] explores the use of freely-available multi-source imagery for accurately identifying forest types using an object-based RF algorithm. The research, conducted in Wuhan, China, used datasets including Sentinel-2A, Sentinel-1A, Shuttle Radar Topographic Mission Digital Elevation Model (DEM), and Landsat-8 images. Results showed that combining Sentinel-2 data with DEM and multi-temporal Landsat-8 imagery significantly improved accuracy, to 82.78%. Zhang et al. (2019) [8] presented the challenges in classifying tropical natural forests due to their intricate structures and the challenging weather conditions. The study focuses on Hainan, China, and utilized multi-temporal synthetic aperture radar (SAR) and optical images from Sentinel-1 and Landsat-8 satellites, respectively. The research proposes a two-stage classification strategy using SVM, incorporating various remote sensing data to identify different types of tropical forests. The approach achieved an overall accuracy of 90% in mapping Hainan's tropical forests. Cheng et al. (2019) [9] introduced an adapted version of dynamic time warping called time-weighted dynamic time warping (TWDTW), for classifying forest types using Sentinel-2 and Landsat-8 time-series images in Southern China. Compared to established ML algorithms like RF and SVM, TWDTW demonstrated superior performance with higher accuracy (93.81%) and stronger agreement (kappa coefficient of 0.93) in mountainous forest-type classification. Hościło et al. (2019) [10] addresses the lack of comprehensive research on regional or national forest status and composition, using Sentinel-2 satellite data. It demonstrated successful classification of forest cover, forest types (broadleaf and coniferous), and identification of tree species in a mountainous area in southern Poland. The RF classifier employed achieved high accuracy in forest/non-forest mapping and forest type classification (98.3% and 94.8% respectively), while the inclusion of topographic data improved accuracy in identifying eight tree species from 75.6% to 81.7% (approach 1) and up to 89.5% for broadleaf and 82% for coniferous species (approach 2). Guo et al. (2021) [11] developed a novel deep fusion uNet model, which utilizes both multi-temporal imagery and the deep uNet model's features. The model demonstrated competitive performance in forest classification, achieving an overall accuracy of 93.30% and a Kappa coefficient of 0.9229 for China's Gaofen-2 satellite data. The model further successfully maps specific plantation species like Chinese pine and *Larix principis*.

In this paper, we explore the utilization of diverse base and advanced (ensemble) ML techniques on multi-temporal remote sensing data for accurate forest type mapping. We further attempt to optimize hyperparameters of a promising base algorithm (SVM) with a view to improving the classification results. Lastly, the paper provides a comprehensive evaluation and comparison of the different methods showcasing their effectiveness in achieving high accuracy and robust performance.

3. Material and Methods

3.1. Data Acquisition

We utilized the forest type mapping dataset [12] from UCI Irvine ML repository for our analysis. This dataset comprises multi-temporal remote sensing data of a forested area in Japan aimed at categorizing different types of forests using spectral data. It was derived from a remote sensing project that employed ASTER satellite imagery to map various forest types based on their spectral properties within the visible-to-near infrared wavelength range. Hence, the primary outcome of this endeavor is a forest type map, which can serve to identify and quantify the ecosystem services offered by the forest, such as carbon storage and erosion protection. The attributes information of the dataset are categorized by forest classes ('s' for 'Sugi' forest, 'h' for 'Hinoki' forest, 'd' for 'Mixed deciduous' forest, and 'o' for 'Other' non-forest land).

3.2. Feature Extraction

We utilized all the 27 features of the dataset for our analysis and included:

- b1 to b9: These are bands of spectral information captured by ASTER imagery, encompassing the green, red, and near-infrared wavelengths, acquired on three different dates (September 26, 2010; March 19, 2011; May 08, 2011).
- pred_minus_obs_S_b1 to pred_minus_obs_S_b9: These values represent the difference between the spectral values predicted through spatial interpolation and the actual spectral values for the 's' class across bands b1 to b9.
- pred_minus_obs_H_b1 to pred_minus_obs_H_b9: Similarly, these values denote the variance between the predicted spectral values obtained via spatial interpolation and the actual spectral values for the 'h' class across bands b1 to b9.

3.3. Model Training

The various ML models were trained by feeding the above features to the models and allowing them learn the relationships between the features and the forest types. We used 70% of the entire data for model training and the remaining 30% for testing. The training process was carried out using basic SVM with no kernel (Linear SVM), SVM with polynomial-3 kernel (Poly-SVM), SVM with Radial Basis Function kernel (RBF-SVM), RBF-SVM with hyperparameter optimization using Grid search (Grid-RBF-SVM) and Bayesian optimization (Bayes-SVM), RF, XGBoost, LightGBM, CatBoost and artificial neural network (ANN).

3.4. Model Evaluation

The trained models were validated on the test data and performance evaluated using metrics such as accuracy, precision, recall, and F1 score.

3.5. SVM Parameter Optimization

For the Grid-RBF-SVM variant, we defined a grid of potential hyperparameters (C: [1e-6, 1e-3, 1, 10, 100, 1e+3, 1e+6] and gamma: [1e-6, 1e-3, 1, 10, 100, 1e+3, 1e+6]) where we varied the regularization parameter (C) and gamma values. The grid search technique exhaustively tested all possible combinations of these hyperparameters within the specified ranges. Also, we utilized Bayesian optimization which is a probabilistic model-based optimization approach to efficiently search for optimal hyperparameters by learning from previous evaluations.

4. Results and Discussion

A comprehensive view of the performance of the different ML models is presented in Table 1 and graphically plotted in Figure 1.

Table 1. Performance of the different ML models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Linear SVM	86.62	86.72	86.62	86.62
Poly-SVM	91.08	91.66	91.08	91.17
RBF-SVM	90.45	91.25	90.45	90.54
Grid-RBF-SVM	93.63	93.81	93.63	93.68
Bayes-SVM	94.27	94.46	94.27	94.32
Random Forest	94.27	94.36	94.27	94.28
XGBoost	93.63	93.88	93.63	93.67
LightGBM	91.72	91.95	91.72	91.76
CatBoost	94.27	94.37	94.27	94.28
ANN	91.08	91.47	91.08	91.09

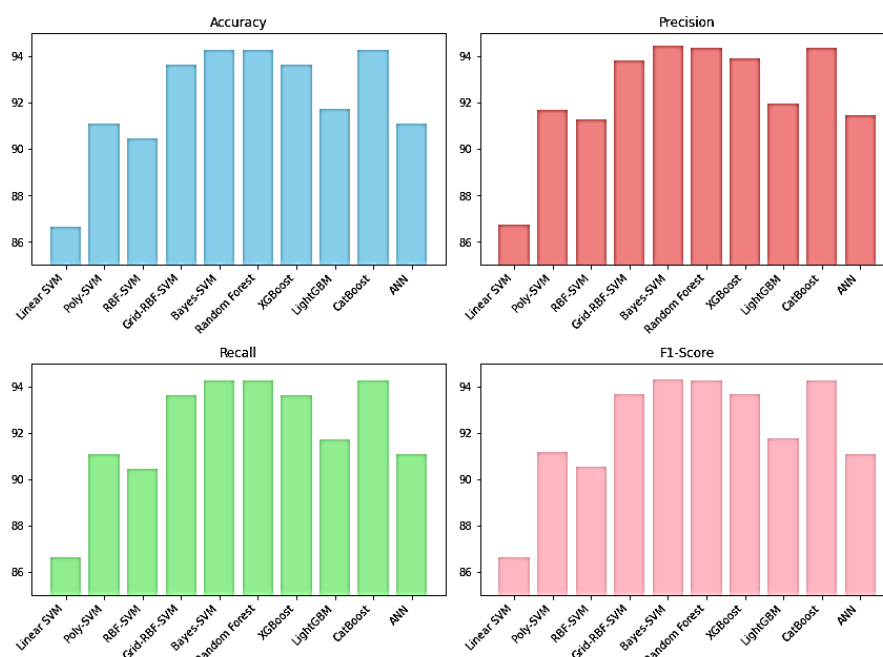


Figure 1. Comparing performance of the different ML models.

From Table 1 and Figure 1, one observation that stands out is the consistent high performance of ensemble methods specifically RF and CatBoost. These models achieved an overall classification accuracy of 94.27%, indicating their efficacy in capturing complex and intricate patterns present in the remote sensing data. Furthermore, they exhibited high Precision, Recall, and F1-Scores above other methods reinforcing their effectiveness.

Another noteworthy finding is the significant performance boost obtained through hyperparameter tuning, particularly with Bayesian optimization. This was evident in the improvement of SVM models, with Bayes-SVM reaching an impressive accuracy of 94.27% and overall best precision and recall of 94.46% and 94.27% respectively which is comparable with those of top-performing ensemble methods. This demonstrates the importance of fine-tuning model hyperparameters to obtain the best possible performance from machine learning algorithms.

Additionally, comparing the performances of Poly-SVM and RBF-SVM against Linear SVM also highlights the advantages of nonlinear SVM variants in capturing complex relationships in the data.

5. Conclusions

In conclusion, this paper demonstrates the effectiveness of several ML models for forest type mapping, with ensemble methods particularly RF and CatBoost yielding exceptional classification performance. Furthermore, hyperparameter tuning with Bayesian optimization was used to enhance SVM model's performance. XGBoost and LightGBM also prove to be dependable alternative models with good performance. These findings highlight the critical role of advanced (ensemble) schemes and hyperparameter optimization in achieving superior results in remote sensing applications. The choice of the final model should be based on a holistic view of performance, computational demands, and practical utilization.

Author Contributions: Conceptualization, Y.I. and U.Y.B.; methodology, Y.I. and A.I.M.; software, Y.I. and A.I.M.; validation, U.Y.B., A.I.M. and Y.I.; investigation, Y.I.; resources, Y.I. and U.Y.B.; data curation, Y.I. and U.Y.B.; writing—original draft preparation, Y.I.; writing—review and editing, Y.I. and A.I.M.; visualization, Y.I. and A.I.M.; supervision, Y.I.; project administration, Y.I. and A.I.M.; funding acquisition, Y.I., U.Y.B. and A.I.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shang, X.; Chisholm, L.A. Classification of Australian native forest species using hyperspectral remote sensing and machine-learning classification algorithms. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *7*, 2481–2489. <https://doi.org/10.1109/JSTARS.2013.2282166>.
2. Boschetti, M.; Boschetti, L.; Oliveri, S.; Casati, L.; Canova, I. Tree species mapping with Airborne hyper-spectral MIVIS data: The Ticino Park study case. *Int. J. Remote Sens.* **2007**, *28*, 1251–1261. <https://doi.org/10.1080/01431160600928542>.
3. Joshi, C.; De Leeuw, J.; van Duren, I.C. Remote sensing and GIS applications for mapping and spatial modelling of invasive species. In Proceedings of the ISPRS, Istanbul, Turkey, 12–23 July 2004; p. B7.
4. Biswas, S.; Huang, Q.; Anand, A.; Mon, M.S.; Arnold, F.-E.; Leimgruber, P. A multi sensor approach to forest type mapping for advancing monitoring of sustainable development goals (SDG) in Myanmar. *Remote Sens.* **2020**, *12*, 3220. <https://doi.org/10.3390/rs12193220>.
5. Ahirwal, J.; Gogoi, A.; Sahoo, U.K. Stability of soil organic carbon pools affected by land use and land cover changes in forests of eastern Himalayan region, India. *Catena* **2022**, *215*, 106308. <https://doi.org/10.1016/j.catena.2022.106308>.
6. Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>.
7. Liu, Y.; Gong, W.; Hu, X.; Gong, J. Forest type identification with random forest using Sentinel-1A, Sentinel-2A, multi-temporal Landsat-8 and DEM data. *Remote Sens.* **2018**, *10*, 946. <https://doi.org/10.3390/rs10060946>.
8. Zhang, L.; Wan, X.; Sun, B. Tropical natural forest classification using time-series Sentinel-1 and Landsat-8 images in Hainan Island. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 6732–6735.
9. Cheng, K.; Wang, J. Forest-type classification using time-weighted dynamic time warping analysis in mountain areas: A case study in southern China. *Forests* **2019**, *10*, 1040. <https://doi.org/10.3390/f10111040>.
10. Hościło, A.; Lewandowska, A. Mapping forest type and tree species on a regional scale using multi-temporal Sentinel-2 data. *Remote Sens.* **2019**, *11*, 929. <https://doi.org/10.3390/rs11080929>.
11. Guo, Y.; Li, Z.; Chen, E.; Zhang, X.; Zhao, L.; Xu, E.; Hou, Y.; Liu, L. A deep fusion unet for mapping forests at tree species levels with multi-temporal high spatial resolution satellite imagery. *Remote Sens.* **2021**, *13*, 3613. <https://doi.org/10.3390/rs13183613>.
12. Johnson, B. Forest Type Mapping. 2015. Available online: <https://archive.ics.uci.edu/dataset/333/forest+type+mapping> (accessed on).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.