iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XI **Month of publication:** November 2024

DOI: https://doi.org/10.22214/ijraset.2024.65465

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

Land Cover Classification in Bamori Range using Machine Learning Algorithms: A Comparative Study

Akshaya Rathore
DFO Guna Forest Division

Abstract: Accurate land cover classification is essential for monitoring the environment and managing natural resources sustainably. In this study, we focused on the Bamori Range in the Guna district of Madhya Pradesh, India, using high-resolution satellite imagery and machine learning to classify land into five categories: bare land, agriculture, fallow cropland, dense forest, and forest. To achieve this, we created a detailed dataset using Sentinel-2 imagery and Dynamic World probabilities, along with feature engineering to improve classification accuracy. We then tested the performance of several models, including Random Forest, Neural Networks, Enhanced Neural Networks, and a hyperparameter-tuned Random Forest, to see which worked best for this task.

I. INTRODUCTION

Recent breakthroughs in machine learning and remote sensing are transforming the way we classify land cover, making mapping more accurate and detailed than ever. By combining high-resolution satellite imagery with advanced algorithms, it's now much easier to analyze different types of land. This study focuses on how various machine learning models perform in classifying land cover in the Bamori Range using these cutting-edge tools. The main goals include building a detailed, high-resolution dataset from satellite images, identifying key features for analysis, testing and comparing different machine learning approaches, and ultimately finding the most effective model. The study also aims to offer insights on how future classification efforts can be improved further.

II. STUDY AREA

The study area includes The Bamori Range, located in the Guna district of Madhya Pradesh, India, is an ecologically significant area requiring detailed land cover analysis for effective management.

III. METHODOLOGY

A. Data Collection

1) Satellite Imagery Acquisition

For this study, we used Planet NICFI satellite imagery from December 2022, focusing on the Bamori Range. This high-resolution imagery was critical for accurately distinguishing between the different land cover types.

2) Training Data Collection

To build the training dataset, we manually categorized the land into five classes using QGIS:
Class 0: Bare Land
Class 1: Cropland
Class 2: Fallow Cropland
Class 3: Dense Forest
Class 4: Forest
We created a shapefile in QGIS and digitized the land cover types by visually interpreting the high-resolution imagery. This hands-on process allowed us to ensure that the training data matched real-world land conditions.

B. Data Preprocessing

Sentinel-2 Preprocessing: We used Sentinel-2 imagery to compute spectral indices, including NDVI, SAVI, and NDWI etc
Feature Combination: To improve classification accuracy, we combined these spectral indices with probabilities from the Google Dynamic World dataset and terrain data. All features were merged into a single composite image to enhance the model's ability to differentiate land cover types..

C. *Label Mapping and Sampling*

We assigned numeric labels to each land cover class within the training polygons. From these labeled areas, we sampled data to generate the training and testing datasets needed for the model. A raster image was created to assign class labels to each pixel, which formed the basis of the machine learning dataset.l.

D. *Data Export and Preparation*

Index Calculation and CSV Export Spectral indices, Dynamic World probabilities, and other feature data were calculated for all the pixels within the training polygons. This enriched dataset, along with the class labels, was exported as a CSV file to prepare it for machine learning model training.

E. *Model Development*

Data Preparation: Before training, the dataset was standardized, and class labels were encoded numerically. The data was split into training (80%) and testing (20%) sets to ensure a fair and reliable evaluation of the model's performance.

Model Selection: We experimented with several machine learning models to find the most accurate and efficient classifier for land cover.

Random Forest: Used as a baseline model due to its robustness and interpretability in handling high-dimensional data.

Enhanced Neural Network: Developed to improve classification performance by capturing non-linear patterns in the data.

Hyperparameter-tuning Random Forest: Implemented by adjusting key parameters for improved accuracy.

F. *Model Training and Evaluation*

We trained the models using the training dataset and evaluated them on the test dataset. To measure performance, we used metrics like accuracy, precision, recall, and F1-score. Confusion matrices provided a detailed breakdown of each model’s ability to classify the five land cover types.

G. *Model Deployment for Temporal Land Classification*

Once the models were developed, we applied them to classify land cover for the same time period across different years. This eliminated the need for retraining, saving time and computational resources. The approach makes it easy to analyze land cover changes over time while maintaining high accuracy and efficiency.

IV. RESULTS AND INTERPRETATION

A. *Model Performance Comparison*

Model	Accuracy (%)
Random Forest (Baseline)	95
Neural Network	92.02
Enhanced Neural Network	93.00
Hyperparameter-tuned RF	96.68

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

B. Confusion Matrices

[[6609 13 106 1 327] [21 7773 57 5 60] [119 175 3527 7 147] [3 24 13 9951 813] [192 39 56 270 16248]]	***(a) Random Forest (Baseline)***
[[6400 20 189 3 444] [17 7706 107 6 80] [181 228 3408 6 152] [6 16 15 9616 1151] [421 43 111 518 15712]]	***(b) Neural Network***
[[6561 18 114 0 363] [18 7736 88 3 71] [189 198 3459 5 124] [2 20 11 9711 1060] [361 42 91 483 15828]]	***(c) Enhanced Neural Network***
[[16345 22 135 2 301] [21 16638 113 4 28] [163 168 16396 5 72] [1 17 22 15971 794] [328 52 139 404 15881]]	***(d) Hyperparameter-tuning Random Forest.***

C. Detailed Analysis of Confusion Matrices

1) Random Forest (Baseline)

Observations:

- High Accuracy in Classifying Cropland (Class 1)
- Significant confusion Between Dense Forest (Class 3) and Forest (Class 4)

2) Neural Network

Observations:

- Increased misclassifications across most classes.
- Significant Misclassifications Between Dense Forest and Forest:
- Misclassifications Involving Bare Land:

3) Enhanced Neural Network

Observations:

- Reduction in misclassifications for most classes.
- Persistent Misclassifications Between Dense Forest and Forest:
- Reduced Misclassifications Involving Bare Land

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

4) *Hyperparameter-tuning Random Forest*

Observations:

- Highest Overall Accuracy Among All Models
- Improved Classification Between Dense Forest and Forest
- Minimal Misclassifications Involving Bare Land:

	precision	recall	f1-score	support
0	0.97	0.97	0.97	16805
1	0.98	0.99	0.99	16804
2	0.98	0.98	0.98	16804
3	0.97	0.95	0.96	16805
4	0.93	0.95	0.94	16804
accuracy			0.97	84022
macro avg	0.97	0.97	0.97	84022
weighted avg	0.97	0.97	0.97	84022

Figure 1. Classification Report of the hyperparameter-tuned Random Forest model

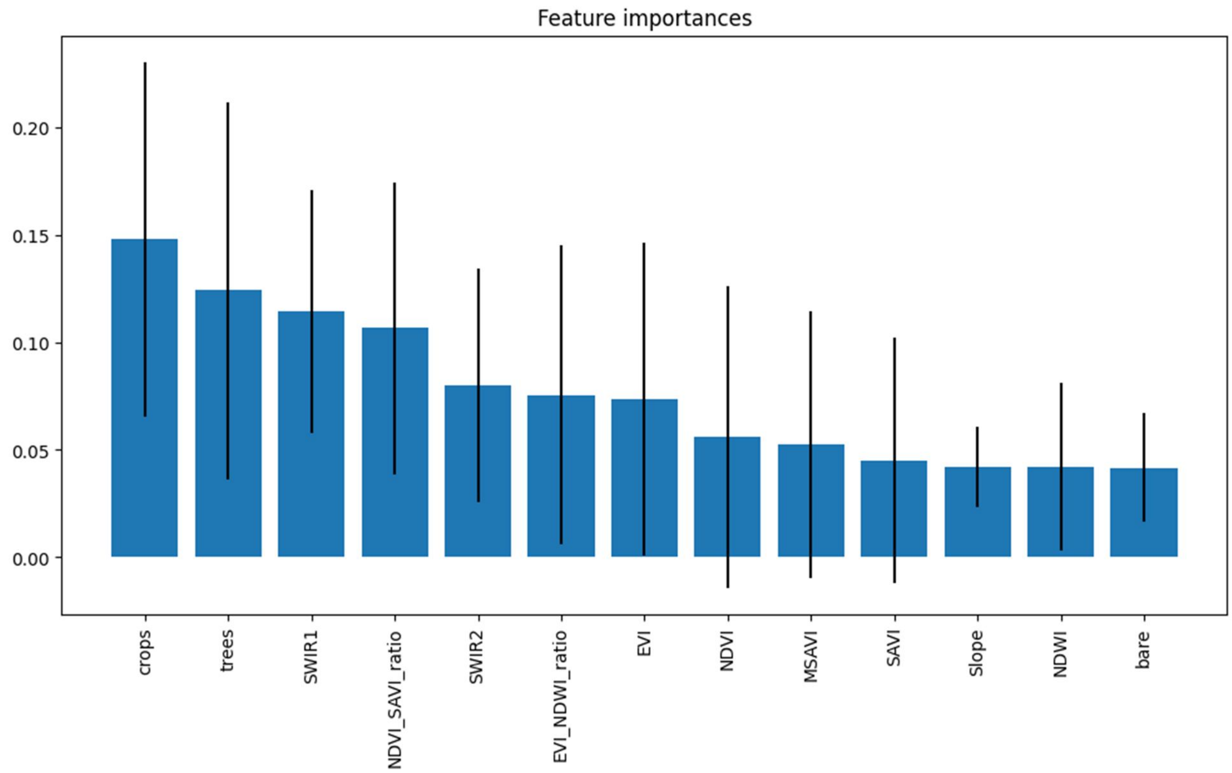


Figure 2. Feature importance plot of the hyperparameter-tuned Random Forest model

Top Features:

- Crops Probability (Dynamic World)
- Trees Probability (Dynamic World)
- SWIR1 Band
- NDVI/SAVI Ratio
- SWIR2 Band

V. DISCUSSION

A. Application

This study's machine learning model transforms land cover classification by eliminating repetitive retraining. Once trained, it can classify satellite images from the same time period across different years, saving time and resources while maintaining accuracy. By relying on robust features like spectral indices and terrain data, the model ensures consistent results, making it ideal for long-term projects like forest management. For example, forest managers can track land cover changes, detect encroachments, or monitor cropland expansion without retraining annually. This efficiency supports faster responses and sustainable practices.

B. Challenges and Limitations

There were a few challenges in this study worth noting. One issue was the persistent misclassification between Dense Forest and Forest. These two land cover types are so similar that it was tough to distinguish them with the available features.Another challenge was the computational cost. Hyperparameter tuning and training the neural networks required a lot of processing power, which meant careful planning to manage resources efficiently. Lastly, the feature set had some limitations. Adding features like texture measurements or structural data from LiDAR could improve the ability to separate similar classes and boost overall accuracy.

VI. FUTURE WORK

To improve the model's performance, a few ideas could be explored. For starters, gradient boosting algorithms like XGBoost or LightGBM might offer better accuracy compared to the current methods. Adding more spectral indices or texture features could also help capture spatial patterns more effectively and reduce confusion between similar classes.

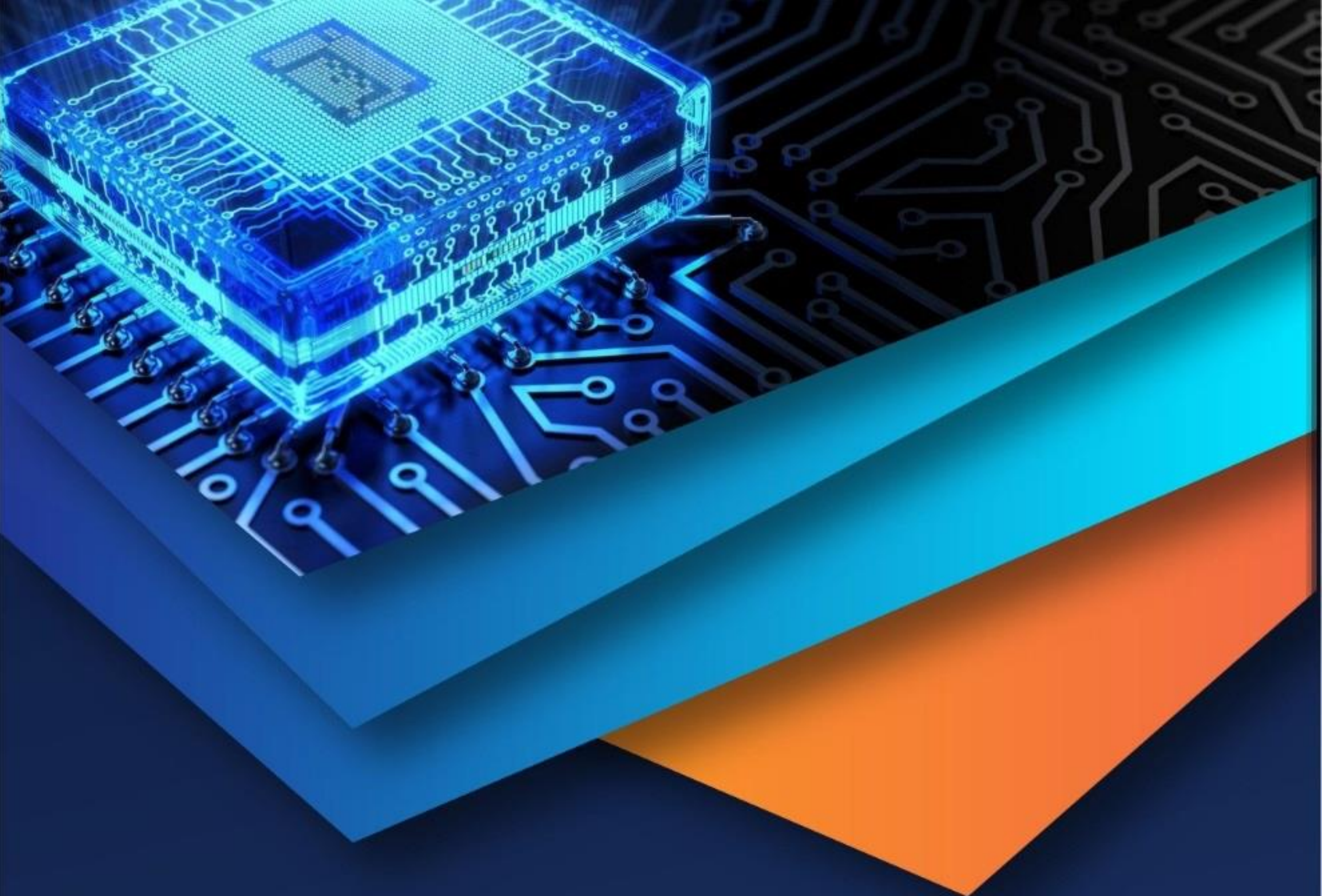
Addressing class imbalances is another priority. Techniques like SMOTE could be used to focus on classes with higher misclassification rates. Additionally, combining models through approaches like stacking or voting classifiers might take advantage of the strengths of each model. Finally, testing the model on data from other time periods or regions would be a good way to see how well it generalizes to new conditions.

VII. CONCLUSION

This study demonstrated the effective use of machine learning algorithms for land cover classification in the Bamori Range. The hyperparameter-tuned Random Forest model achieved the highest accuracy, highlighting the importance of model tuning and feature selection. Integrating Dynamic World probabilities and spectral indices significantly enhanced classification performance. Future work will focus on exploring advanced algorithms and feature engineering to further improve accuracy.With the help of this machine learning model , we can easily get the land classification for the same time period for any year and we don't need to use the training classification again and again . So, with these models we can directly get the land classification with this high accuracy and escape the cumbersome training activity for supervised classification.

REFERENCES

[1] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324
[2] Planet Labs Inc. (2022). Planet NICFI Satellite Imagery. Retrieved from https://www.planet.com/nicfi/
[3] Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18-27. https://doi.org/10.1016/j.rse.2017.06.031
[4] Wang, L., Sousa, W. P., & Gong, P. (2004). Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery. International Journal of Remote Sensing, 25(24), 5655-5668. https://doi.org/10.1080/01431160410001726021
[5] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.
[6] Dynamic World (2022). A near real-time global land use and land cover map powered by Google Earth Engine. Retrieved from https://dynamicworld.app
[7] Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment, 8(2), 127-150. https://doi.org/10.1016/0034-4257(79)90013-0
[8] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). https://doi.org/10.1145/2939672.2939785
[9] Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25(3), 295-309. https://doi.org/10.1016/0034-4257(88)90106-X
[10] Pal, M. (2005). Random forest classifier for remote sensing classification. International Journal of Remote Sensing, 26(1), 217-222. https://doi.org/10.1080/01431160412331269698
[11] Van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. Computing in Science & Engineering, 13(2), 22-30. https://doi.org/10.1109/MCSE.2011.37
[12] Zhu, X., Gerber, J. S., Carlson, K. M., & West, P. C. (2019). Spatially explicit global cropland sustainability metrics. Nature Sustainability, 2(6), 460-468. https://doi.org/10.1038/s41893-019-0292-6



10.22214/IJRASET

INDEX COPERNICUS 45.98

ISRA JIF IMPACT FACTOR: 7.129

SJIFactor TOGETHER WE REACH THE GOAL IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY