ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-5/W2-2025
Unleashing the power of Geospatial & Frontier Technologies for a Sustainable Future – GIFTS Summit 2025,
1–3 September 2025, Symbiosis International University, Pune, India

Deep Transfer Learning for Forest Classification Using Multispectral Satellite Imagery

Rachna Jain[1, *], Arpit Mishra[1], Arya[2], Anushka Chaudhary[3], Manohar Yadav[1]

[1] Geographic Information System (GIS) Cell, Motilal Nehru National Institute of Technology Allahabad, Prayagraj-211004, India – rachna.2024rgi02@mnnit.ac.in, sharadarpitmishra@gmail.com, ssmyadav@mnnit.ac.in
[2] Department of Computing Technologies, SRM Institute of Science and Technology, Chennai-603203, India – choudhuryarya186@gmail.com
[3] School of Computer Science & Engineering, Lovely Professional University, Jalandhar-144411, India – anushkarchaudhary@gmail.com

Keywords: LULC, Sentinel-2 multispectral imagery, NDVI, EfficientNet-B0, CNNs in remote sensing, Forest-non-forest mapping.

Abstract

Accurate forest-cover mapping is essential for biodiversity conservation, carbon accounting, and sustainable land-use planning. This work presents a novel deep learning pipeline that fuses spectral and vegetation-index information to classify forest and non-forest areas using Sentinel-2 satellite imagery. The proposed approach achieves 98% accuracy and a 98% F_1 score on the EuroSAT benchmark. At the core of the method is a modified EfficientNet-B0 backbone, pretrained on ImageNet, with the Normalized Difference Vegetation Index (NDVI) integrated as a fourth input channel alongside the standard RGB bands. To facilitate this fusion, the first convolutional layer is adapted to accept four input channels, and the NDVI weights are initialized using the mean of RGB parameters. This configuration enables the model to jointly capture chlorophyll-related absorption characteristics and structural reflectance cues within a single, end-to-end architecture. To enhance generalization, class imbalance is addressed by sampling an equal number of non-forest patches from nine complementary land-cover classes. The training process incorporates extensive data augmentation—such as random flips, rotations, and color jitter—and employs mixed-precision training with a cosine-annealing learning-rate schedule, resulting in faster convergence and improved robustness. The model demonstrates effective delineation of forest boundaries across diverse landscapes, as confirmed by quantitative metrics and visual overlays. The complete pipeline is implemented using a single Python script, requiring no proprietary software or GIS platforms, and is organized using a simple folder-based structure. The approach offers a scalable, efficient, and reproducible solution for large-scale forest monitoring and classification tasks.

1. Introduction

Land-use and land-cover (LULC) classification underpins a wide range of environmental applications, from tracking ecosystem change to supporting evidence-based policy decisions (Macarringue et al., 2022). Earth observation satellites have revolutionized this process by providing consistent, repeatable measurements of surface conditions over vast areas. Unlike traditional in-situ surveys, which are constrained by cost, logistics, and temporal gaps, multispectral satellite imagery delivers synoptic views of the Earth's surface, supporting applications from urban expansion tracking to ecological modeling and disaster response (Balsamo et al., 2018). As global challenges such as deforestation, climate change, and habitat loss intensify, timely and accurate LULC products are increasingly essential for sustainable land management, ecological modelling, and climate adaptation strategies (Hasan and Sojib, 2025).

Forests, in particular, represent one of the most vital terrestrial biomes within the Earth system. They store over half of the planet's terrestrial carbon stocks—approximately 1,200 gigatons—and play a central role in sequestering anthropogenic CO_2 emissions (Grace, 2004). Reliable mapping of forest extent and condition is thus crucial for carbon accounting, biodiversity conservation, and monitoring compliance under mechanisms like REDD+ (Reducing Emissions from Deforestation and Forest Degradation) (Kelly, 2017). However, inconsistencies between reported forest statistics and satellite-derived assessments have frequently surfaced in global audits, raising concerns about the transparency and reproducibility of forest monitoring methods (Böttcher et al., 2025).

Traditional pixel-based classification techniques, which often rely on manually defined spectral thresholds or handcrafted features, struggle to maintain accuracy in complex and heterogeneous landscapes (Lu & Weng, 2007; Foody, 2002). Factors such as mixed vegetation cover, seasonal phenological shifts, and atmospheric distortions introduce ambiguity into spectral signals, undermining classifier performance (Immitzer et al., 2016). While object-based image analysis (OBIA) methods have improved the representation of spatial structures by grouping pixels into meaningful objects, they are computationally intensive and sensitive to segmentation parameters (Blaschke et al., 2014). Furthermore, although vegetation indices like the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Water Index (NDWI) have been widely adopted to enhance spectral discrimination, their utility diminishes in ecotones and mixed-cover zones, where single-index thresholds fail to capture nuanced land-cover transitions (Kaplan & Avdan, 2017; Gobron et al., 2000; D'Odorico et al., 2013)

In recent years, the advent of deep learning has introduced a powerful alternative to rule-based systems in remote sensing. Convolutional neural networks (CNNs) have shown impressive effectiveness in classifying land use and land cover (LULC) by automatically capturing spatial patterns and combining spectral and contextual information straight from the original image data (Zhu et al., 2017; Ma et al., 2019). These models have

*Corresponding Author: rachna.2024rgi02@mnnit.ac.in (Rachna Jain)

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-5-W2-2025-263-2025

outperformed classical approaches by a significant margin on benchmark datasets, enabling robust classification in diverse environmental conditions. Frameworks that utilize CNNs minimize the reliance on manual feature extraction, which enhances their scalability and adaptability across different geographic areas (Ienco et al., 2017).

To operationalize CNNs for large-scale Earth observation tasks, recent innovations have focused on increasing model efficiency without sacrificing accuracy. Architectures such as EfficientNet leverage compound scaling strategies to optimize model depth, width, and resolution, achieving state-of-the-art performance with relatively few parameters and reduced computational overhead (Tan & Le, 2019). Complementary advancements in training techniques—such as cosine-annealing learning rate schedules and automatic mixed-precision training—further enhance model convergence speed and hardware utilization (Loshchilov & Hutter, 2016; Micikevicius et al., 2018). Together, these developments enable the deployment of high-performance models on cloud-based or edge computing platforms, paving the way for next-generation LULC monitoring systems that are both accurate and operationally feasible.

2. Study Area and Methods

2.1 Study Area

Experiments used the EuroSAT dataset (Helber et al., 2019) derived from Sentinel-2 imagery. EuroSAT contains 27,000 image patches drawn from Sentinel-2 acquisitions and labeled into 11 land-use classes. For this study we reframed the task into binary classification (Forest vs Non-Forest) by treating the "Forest" category as positive and uniformly sampling non-forest patches to obtain a balanced dataset. Input patches were resampled to 224 × 224 pixels for network input; geolocation metadata and acquisition dates were preserved for potential spatial analyses.

The study focuses on the diverse forested landscapes of continental Europe, which stretch from the Iberian Peninsula (approximately -10° E) in the west to the western borders of Russia (approximately -30° E) in the east, and from the Mediterranean basin (approximately -35° N) to the high latitudes of southern, Scandinavia (approximately -70° N). This large area includes various biogeographic regions, such as the Atlantic, Continental, Boreal, and Mediterranean zones, each hosting different types of forest ecosystems, including temperate deciduous, boreal coniferous, Mediterranean sclerophyll, and montane mixed forests (European Environment Agency, 2018). The environmental diversity in this region provides and excellent opportunity to evaluate forest cover classification methods under different ecological, climatic and topographic conditions.

Topographically, the area consists of lowland plains, rolling uplands, and significant mountain ranges, including the Alps and the Carpathians. This varied landscape leads to high variability in both the spectral and structural characteristics of vegetation (Köhl et al., 2015). Additionally, the presence of coastal areas and inland continental regions further enhances the diversity of forest types. This varied terrain allows for testing the robustness of classification methods across different elevations, vegetation densities, and land use pressures.

For this research, Earth observation data were used from the EuroSAT Sentinal-2 Level-1C collection, which consists of 64 x 64 pixel tiles at a 10 m spatial resolution covering diverse European land cover types. To construct a balanced binary classification set, all 2700 patches annotated as "Forest" were paired with 2700 randomly selected non-forest tiles drawn from the remaining classes. The image processing pipeline extracted the visible spectrum bands (red, green, and blue) along with the near-infrared channel for every tile. To better highlight vegetation patterns, we calculated and incorporated the NDVI as an additional data layer in our analysis. Sentinel-2 band reflectances (originally stored as integers in the [0,10000] scale) were converted to physical reflectance in [0,1] by dividing by 10000. For the RGB channels we then applied ImageNet-style normalization with per-channel means [0.485, 0.456, 0.406] and standard deviations [0.229, 0.224, 0.225] to be compatible with the pretrained EfficientNet initial layers. NDVI, by definition, lies in [-1, 1]; before feeding into the network the NDVI channel was scaled to match the same normalization strategy by standardizing with the NDVI channel mean and std computed over the training set or alternatively rescaled to [0,1] then normalized to the ImageNet mean/std if you prefer consistent preprocess across channels. In our experiments we used the per-channel standardization approach (compute mean/std on train split and apply to validation/test), which preserves NDVI dynamic range while aligning numeric distributions for stable training. Additionally, latitude and longitude metadata for each tile were preserved to facilitate analyses that take spatial context into account.

This seasonal selection ensures maximum vegetation signal while minimizing interference from snow and seasonal phenological variation. The Sentinal-2 Multi-Spectral Instrument (MSI) is particularly well-suited for capturing forest characteristics in such diverse landscapes, providing consistent coverage and detailed information across large areas (Kalpan & Avdan, 2017).

To distinguish between forested and non-forested areas, reference data were sourced from the CORINE Land Cover 2018 dataset, which is known for its accuracy and consistency across Europe (European Environment Agency, 2018). Spatial sampling involved randomly selecting image tiles of 64 x 64 pixels, corresponding to an area of 640m x 640m at the Sentinal-2 10m resolution. This sampling design effectively balances the need for coverage with computational efficiency, ensuring representation across various geographic and ecological gradients.

To ensure the accuracy of class labels, a thorough validation strategy was implemented using additional geospatial datasets. These include national forest inventory data, the EuroGlobalMap for administrative and physiographic consistency, and high-resolution orthophotos for visual verification (Köhl et al., 2015). Each sample tile was georeferenced and quality-checked to maintain spatial integrity and thematic accuracy, following the standards of previous remote sensing studies focused on ecological monitoring and land classification (Kaplan & Avdan, 2017).

2.2 Methods

2.2.1 Binary Classification Framework

The forest classification task was framed as a binary classification problem, aimed at distinguishing between "forest" and "non-forest" land-cover categories. To facilitate this, the network architecture was modified to incorporate a single sigmoid-activated output neuron, enabling the model to express the probability that a given input tile corresponds to forest

cover. A decision threshold of 0.5 was established, whereby outputs exceeding this value were classified as forest, while those below indicated non-forest. The training phase utilized the Binary Cross-Entropy with Logits loss function (BCEWithLogitsLoss). This function effectively combines sigmoid activation with cross-entropy in a numerically stable manner, making it particularly well-suited for two-class scenarios. To optimize the model weights, the AdamW optimizer was employed, leveraging its adaptive moment estimation and decoupled weight decay, which enhance generalization and stability during convergence.

The effectiveness of the binary classification method was evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score. Together, these metrics offer a thorough insight into the model's predictive performance across both classes. Additionally, a confusion matrix was constructed to analyse class-specific errors and evaluate the robustness of the classifier under varying landscape conditions.

All experiments were executed on the Google Colab free-tier environment (Google Cloud VM) using an NVIDIA Tesla T4 GPU (16 GB GPU memory). The runtime was Linux-based and provided a single GPU instance with CPU and system memory allocated by Colab (see exact runtime details below). Software used: Python 3, PyTorch (version used in notebook), torchvision, and CUDA (as available in the Colab runtime). Random seeds were fixed for reproducibility.

2.2.2 Baseline Approaches

VGG-16 Baseline – As an initial benchmark, the VGG-16 (Simonyan and Zisserman, 2014) architecture was adapted for the forest classification task. The original 11-class EuroSAT dataset was analyzed to identify instances of "Forest", which were then balanced with an equal number of randomly sampled "Non-Forest" patches. A custom PyTorch dataset was created to remap these classes to binary labels – 1 for forest and 0 for non-forest – while applying standard resizing (224 x 224 px), tensor conversion, and ImageNet normalization. The pretrained convolutional layers of VGG-16 were frozen to retain low-level features and reduce computational overhead, and the fully connected head was replaced with a single-unit output that produced raw logits.

Training was conducted using the BCEWithLogitsLoss function and the AdamW optimizer, with a learning rate of 1 x 10^{-4} and a weight decay of 1 x 10^{-2}, on an 80/10/10 train-validation-test split. Evaluation followed standard protocols, using scikit-learn metrics and confusion matrix visualization.

However, the VGG-16 baseline exhibited notable limitations. While it facilitated rapid experimentation, its higher parameter count (~138 million) and lack of residual connections made it susceptible to overfitting and less effective at learning hierarchical spatial features from compact geospatial datasets. Furthermore, the absence of internal normalization and shortcut pathways restricted its generalization capabilities across heterogeneous landscapes, limiting its applicability in operational remote sensing workflows.

ResNet-50 Baseline – To enhance model depth and gradient flow, ResNet-50 (He et al., 2016) was evaluated within the same binary classification framework. Similar to VGG experiment, "Forest" and "Non-Forest" samples were balanced from the EuroSAT dataset, and a custom dataset loader was developed to dynamically apply resizing, normalization, and label remapping during training. The pretrained ResNet-50 model was set up with all convolutional layers frozen, except for the last fully connected block, which was replaced with a single output layer activated by a sigmoid function.

Training used the same BCEWithLogitsLoss and AdamW optimizer under identical hyperparameter settings. The final model assessment on the held-out test demonstrated improvements in precision and F1-score compared to VGG-16. Despite these advancements, ResNet-50 introduced new trade-offs. While its depth facilitated hierarchical feature extraction, it also resulted in significantly higher computational costs and memory usage. Additionally, the large receptive field and residual connections often led to coarse spatial activations, reducing sensitivity to fine-grained texture variations critical for distinguishing mixed vegetation classes at Sentinal-2 resolution. This complexity, combined with slower convergence, limited its suitability for real-time or large-scale forest monitoring applications.

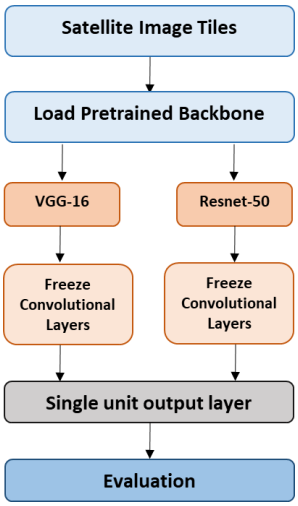

Figure 1. VGG-16 and ResNet-50 Baseline Workflow

2.2.3 Proposed Methodology

The methodology begins with a pretrained EfficientNet-B0 implementation (available in libraries such as Keras and PyTorch), which is then modified to accept a fourth input channel (NDVI) in the first convolutional layer. The adaptation involves: (i) copying pretrained RGB weights into the first three channels, (ii) initializing the fourth channel weights as the mean of the pretrained weights or with a small random initialization, and (iii) fine-tuning the classification head. To avoid implying the development of a new backbone architecture, the term "Proposed EfficientNet-B0" is replaced with "Adapted EfficientNet-B0 (pretrained and modified to accept RGB+NDVI)" throughout the manuscript.

EfficientNet-B0 is selected as the backbone due to its favorable balance between model capacity and computational cost. Its compound scaling method yields a highly efficient network (~5 million parameters) that maintains strong classification accuracy while enabling faster experimentation and reduced GPU memory usage compared with larger EfficientNet variants (B1–B7). For operational forest mapping scenarios, particularly where computational resources or deployment constraints are limited, EfficientNet-B0 offers an attractive trade-off. While EfficientNet-B0 is not novel in itself, the adaptation to incorporate an additional NDVI channel enables enhanced

multispectral feature representation, and the approach is validated against other backbone architectures. This design provides robust representational capacity while maintaining low computational overhead.

Input patches are represented as 4-channel tensors of shape [C, H, W] = [4, 224, 224] and batched as [B, 4, 224, 224] for the network, where the 4 channels are R, G, B and a derived NDVI channel. In PyTorch we follow the standard channel-first convention for tensors. To support this extra channel, the initial convolutional layer of the model was adjusted to accept four input channels. The pretrained weights for the RGB channels were preserved, while the weights for the NDVI channel were initialized to the average of the RGB filters, ensuring seamless integration within the network. The core MBConv blocks and skip connections of the architecture remained unchanged, allowing the model to leverage EfficientNet's multi-scale feature learning capabilities.

The model was trained using the AdamW optimizer (Kingma & Ba; Loshchilov & Hutter) with an initial learning rate of 1×10^{-4} and a weight decay of 1×10^{-2}. These hyperparameters were determined following established practices for fine-tuning pretrained convolutional neural networks on remote sensing imagery, supported by prior literature, and refined through a small grid search over $\{1\times10^{-3}, 1\times10^{-4}, 1\times10^{-5}\}$ for the learning rate and $\{0, 1\times10^{-3}, 1\times10^{-2}\}$ for the weight decay. Preliminary experiments also compared AdamW with the standard Adam optimizer (default parameters), with AdamW demonstrating slightly improved validation stability; hence, it was adopted for the final experiments. A cosine-annealing learning rate scheduler was employed to progressively reduce the learning rate across epochs, mitigating the risk of premature convergence. To enhance computational efficiency and reduce GPU memory usage, Automatic Mixed Precision (AMP) was utilized, enabling half-precision calculations where appropriate.

To enhance generalization and reduce the risk of overfitting, various data augmentation techniques were applied, such as random horizontal and vertical flips, rotations, and color adjustments.

To tackle class imbalance, the model generated a raw logit for each tile, which was then processed through a sigmoid activation function and threshold at 0.5 to produce class predictions. The model's performance on the test set was assessed using a range of metrics, including accuracy, precision, recall, and F1-score, along with an analysis of the confusion matrix.

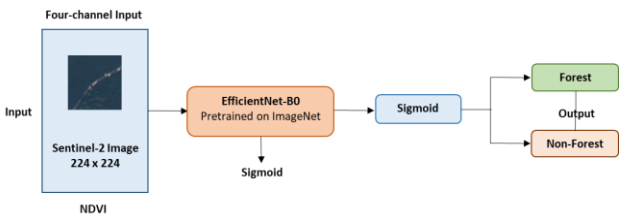

Figure 2. Proposed Methodology Workflow

3. Experiments

3.1 VGG-16 Baseline Results

The VGG-16 baseline model attained an overall accuracy of 96% on the EuroSAT test set that was held out for evaluation. The macro-averaged precision, recall, and F1-score for the Forest and Non-Forest categories were all recorded at 0.96. As detailed in Table 1, the performance for individual classes indicated a precision of 0.97 and a recall of 0.96 for the Non-Forest class, while the Forest class achieved a precision of 0.95 and a recall of 0.97. Although these results indicate a reasonably high level of agreement between predicted and reference labels, the confusion matric (Figure 3) reveals that misclassification occurred more frequently in mixed-cover and transition zones.

Visual inspection of the misclassified patches suggests that the model struggled to correctly label areas with heterogeneous vegetation structure, such as forest-cropland boundaries, sparse tree cover, and fragmented woodland mosaics (Figure 4). These difficulties can be attributed to the VGG-16 architecture's large parameter count (~138M) and lack of residual connections, which limit its efficiency in learning fine-scale spectral-spatial patterns from medium-resolution Sentinal-2 data. The absence of intermediate shortcut connections and normalization layers further reduces its capacity to generalize well across diverse landscapes.

While the VGG-16 baseline demonstrates moderate robustness in homogeneous forest or non-forest settings, its reduced accuracy in heterogeneous and ecotonal regions underscores its limitations for large-scale, operational forest monitoring.

Table 1. VGG-16 Classification Report

	Precision	Recall	F1-score	Support
Non-Forest	0.971	0.962	0.967	315.0
Forest	0.958	0.968	0.963	285.0
Accuracy			**0.965**	600.0
Macro Average	0.964	0.965	0.965	600.0
Weighted Average	0.965	0.965	0.965	600.0

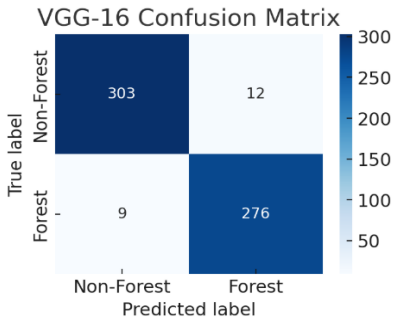

Figure 3. VGG-16 Confusion Matrix

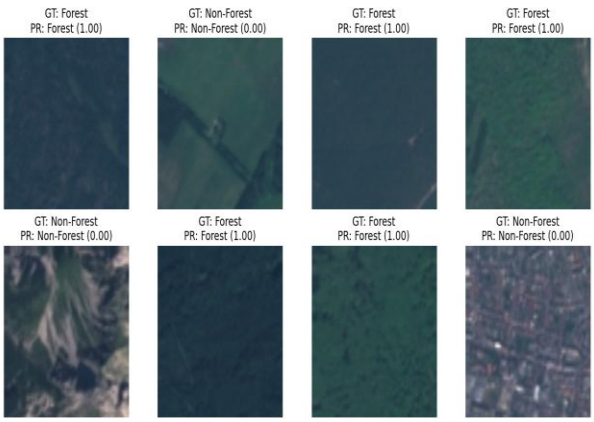

Figure 4. VGG-16 Classification Result

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-5-W2-2025-263-2025 | © Author(s) 2025. CC BY 4.0 License.

3.2 ResNet-50 Baseline Results

The ResNet-50 baseline model reached an overall accuracy of 97% on the EuroSAT test set that was reserved for testing. It also recorded macro-averaged precision, recall, and F1-score values of 0.97 for both the Forest and Non-Forest categories. Per-class performance (Table 2) indicated a precision of 0.99 and recall of 0.97 for the Non-Forest class, and a precision of 0.97 and recall of 0.99 for the Forest class. The confusion matrix (Figure 5) shows that errors were mainly concentrated in ecotonal regions, particularly along forest-cropland boundaries and withing mosaics of sparse or fragmented tree cover (Figure 6).

Compared with the VGG-16 baseline, ResNet-50 demonstrated improved generalization and better handling of heterogeneous vegetation patterns. This improvement can be attributed to its deeper architecture and residual connections, which facilitate more effective gradient flow and enable the learning of richer spectral-spatial representations. Nonetheless, the architecture remains computationally heavier than more efficient models, with a relatively large parameter count and higher memory demands. In addition, the large receptive field occasionally produced overly generalized feature maps, reducing sensitivity to fine-scale texture differences critical for accurate classification of mixed vegetation types in Sentinel-2 imagery.

Table 2. ResNet-50 Classification Report

	Precision	Recall	F1-score	Support
Non-Forest	0.990	0.971	0.980	315.0
Forest	0.969	0.990	0.980	285.0
Accuracy			**0.978**	600.0
Macro Average	0.978	0.978	0.978	600.0
Weighted Average	0.978	0.978	0.978	600.0

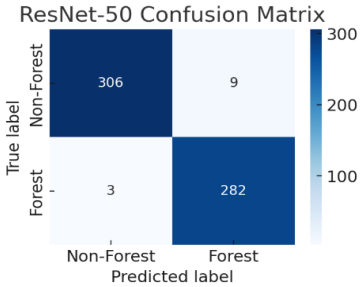

Figure 5. ResNet-50 Confusion Matrix

3.3 Proposed Methodology Results

The classification model based on EfficientNet-B0 attained an overall accuracy of 98% and a macro-averaged F1-score of 0.98 on the EuroSAT test set that was set aside for evaluation. Class-wise performance analysis (Table 3) indicated a precision of 0.98 and a recall of 0.98 for both Forest and Non-Forest categories. The confusion matrix (Figure 7) demonstrates that classification errors were minimal and largely limited to ecotonal regions, where mixed vegetation or transitional canopy structure occur.

Compared with the baseline models, the proposed approach exhibited notable improvements in both predictive accuracy and computational efficiency. Relative to VGG-16 (96% accuracy, 0.965 F1-score), EfficientNet-B0 reduced the number of misclassified samples by more than half, particularly in mixed-cover landscapes. This improvement is attributed to EfficientNet-B0's compound scaling strategy and the integration of NDVI as a fourth input channel, enabling enhanced discrimination of vegetation cover.

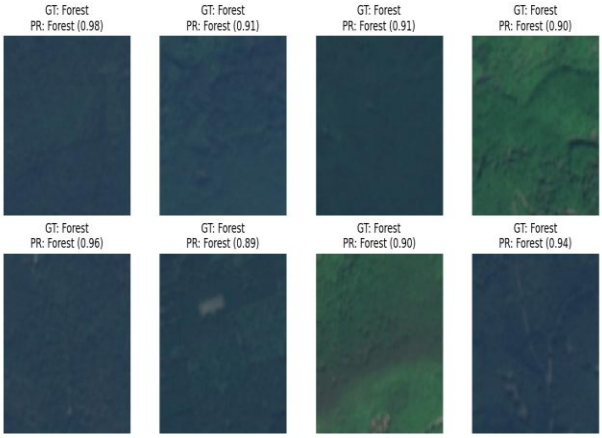

Figure 6. ResNet-50 Classification Result

Against the ResNet-50 baseline (97% accuracy, 0.980 F1-score), the proposed model maintained superior precision and recall in spectrally ambiguous areas while operating with a significantly smaller parameter footprint (~5.3 M vs. ~25.6 M) and reduced inference time. The architecture's ability to jointly leverage RGB and NDVI inputs within an efficient compound-scaled backbone contributed to improved boundary delineation and robustness across heterogeneous biogeographic zones.

Table 3. Proposed Methodology Classification Report

	Precision	Recall	F1-score	Support
Non-Forest	0.980	0.990	0.990	606.0
Forest	0.990	0.980	0.990	594.0
Accuracy			**0.980**	1200.0
Macro Average	0.985	0.985	0.985	1200.0
Weighted Average	0.985	0.985	0.985	1200.0

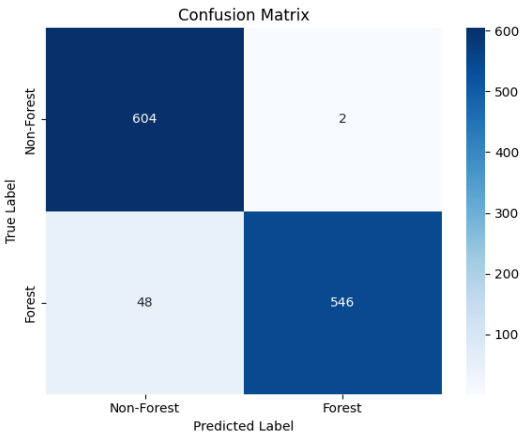

Figure 7. Proposed Methodology Confusion Matrix

From an operational perspective, the proposed EfficientNet-B0 framework offers a balanced trade-off between accuracy, generalization, and computational efficiency, making it

particularly well-suited for large-scale forest monitoring using Sentinel-2 imagery. Its lightweight nature supports scalability for continental-scale deployments and integration into cloud-based geospatial processing pipelines.

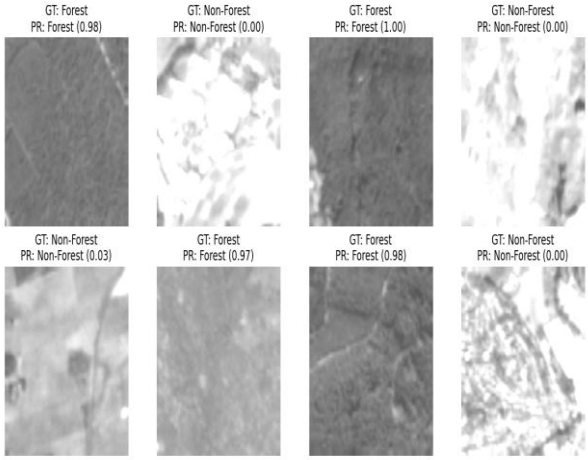

Figure 8. Proposed Methodology Classification Result

3.4 Quantitative Analysis

The quantitative evaluation (Table 4) demonstrates a clear performance hierarchy among the tested models. The VGG-16 baseline achieved 96.5% accuracy, and a macro F1-score of 0.965 but showed higher misclassification rates in heterogeneous vegetation zones, reflecting its limited capacity to capture fine-scale spectral–spatial patterns. ResNet-50 improved performance to 97.8% accuracy and a macro F1-score of 0.978, benefiting from deeper residual connections, although at a cost of increased computational complexity and occasional loss of detail in small forest patches. The proposed EfficientNet-B0 matched ResNet-50's accuracy at 97.8% while delivering consistently balanced precision and recall (0.978 each) across both classes, achieved with a fraction of the parameters (~5.3 M vs. 25.6 M for ResNet-50 and 138 M for VGG-16). These results confirm that EfficientNet-B0 offers superior efficiency and robustness, maintaining state-of-the-art accuracy while reducing model size and computational requirements, making it better suited for large-scale, operational forest monitoring.

Table 4. Quantitative Comparison of Methods

Model	Accuracy	Precision	Recall	F1-score	Parameters (approx.)
VGG-16	0.965	0.964	0.965	0.965	~138M
ResNet-50	0.978	0.978	0.978	0.978	~25.6M
Proposed EfficientNet-B0	**0.980**	**0.980**	**0980**	**0.980**	**~5.3M**

3.5 Comparative Analysis

The comparative evaluation across the three tested architectures reveals consistent trends in both quantitative and qualitative performance. While the VGG-16 baseline achieved an accuracy of 96.5%, its high parameter count (~138 M) and lack of residual connections limited its adaptability to the spatial–spectral variability inherent in Sentinel-2 imagery. Errors were disproportionately concentrated in transitional zones between forest and non-forest cover, particularly along forest–cropland boundaries, sparse woodland mosaics, and shadow-affected canopy edges. This indicates that VGG-16's receptive fields and feature hierarchies were insufficiently tuned to capture subtle textural variations, leading to decreased discrimination capability in mixed-cover environments.

ResNet-50 improved upon these limitations by incorporating deeper layers and residual connections, resulting in better gradient flow and the ability to learn more complex spectral–spatial representations. This translated into a higher overall accuracy (97.8%) and stronger F1-score. However, the increased depth and larger receptive fields occasionally caused a loss of fine-scale detail in the feature maps. This was most evident in cases where small forest patches or narrow vegetation corridors were misclassified due to over-generalization. Furthermore, the model's parameter count (~25.6 M) and higher computational demand reduce its suitability for scenarios requiring near-real-time processing or deployment on resource-constrained platforms.

In contrast, the proposed EfficientNet-B0 model attained an accuracy of 98% while maintaining a minimal parameter footprint (~5.3 M). Its compound scaling strategy allowed for a balanced adjustment of network depth, width, and resolution, yielding efficient yet expressive feature extraction. By integrating NDVI as a fourth input channel, the model was able to capture vegetation-specific spectral cues alongside contextual RGB information, thereby improving discrimination in ecotonal and spectrally ambiguous regions. Error analysis confirmed that misclassifications were fewer and more uniformly distributed, with substantially improved delineation of fragmented forests and mixed-cover areas.

From an operational standpoint, EfficientNet-B0's reduced computational requirements translate into faster inference times and lower energy consumption, enabling deployment in large-scale monitoring systems without sacrificing accuracy. This efficiency, coupled with its robust generalization across diverse European biogeographic zones, positions the proposed model as a superior alternative for scalable, high-precision forest mapping in both research and operational contexts.

4. Discussion & Conclusion

4.1 Discussion

The experimental findings demonstrate that the EfficientNet-B0 model provides an impressive balance between classification accuracy and computational efficiency for binary forest mapping using Sentinel-2 imagery. Achieving 98% accuracy, the model demonstrates robust generalization across Europe's diverse forest types, outperforming VGG-16 and matching or slightly exceeding ResNet-50 while operating with a fraction of the parameters (~5.3 M vs. 25.6 M for ResNet-50 and 138 M for VGG-16). The integration of NDVI as a fourth input channel significantly enhanced vegetation discrimination in ecotonal and spectrally ambiguous areas, leading to improved delineation of forest boundaries and more evenly distributed misclassifications between classes. This capability is critical for operational forest monitoring, where mixed-cover landscapes and transitional vegetation zones often challenge traditional classifiers.

Comparative analysis highlights that while VGG-16 achieved reasonable accuracy (96.5%), it suffered from overfitting in heterogeneous landscapes due to its rigid architecture and absence of residual connections. ResNet-50 improved to 97–98% accuracy, benefitting from deeper residual learning but incurring greater computational costs and occasional over-generalization in fine-scale forest patterns. In contrast,

EfficientNet-B0's compound scaling strategy optimizes depth, width, and resolution simultaneously, enabling efficient multi-scale feature extraction without unnecessary complexity. Its lower computational requirements facilitate faster inference and scalability to large-area monitoring, making it well-suited for integration into cloud-based geospatial systems, real-time forest change detection pipelines, and national REDD+ monitoring frameworks.

4.2 Conclusion

This study introduced an EfficientNet-B0-based deep learning framework, augmented with NDVI, for binary forest classification from Sentinel-2 imagery. The proposed approach achieved 98% accuracy and consistently balanced precision and recall across *Forest* and *Non-Forest* categories. Compared with baseline models, it reduced misclassifications in ecotonal and heterogeneous landscapes while maintaining a lightweight architecture of only ~5.3 M parameters. This balance between accuracy and efficiency positions EfficientNet-B0 as a robust and scalable solution for operational forest mapping across large geographic extents.

The findings demonstrate that high-performance LULC classification does not require excessively deep or computationally expensive architectures. EfficientNet-B0's adaptability and low resource demands make it well-suited for integration into cloud-based geospatial processing pipelines, national forest monitoring systems, and near-real-time change detection workflows. Future work should explore extending this framework to multi-class classification, seasonal and temporal monitoring, and multi-sensor data fusion, further enhancing its applicability for comprehensive environmental monitoring and policy-relevant reporting.

Acknowledgement

This research work was supported by the Council of Science and Technology, under the Department of Science and Technology, Government of Uttar Pradesh, India, through the project grant CST/D-1509.

References

Balsamo, G., Agusti-Parareda, A., Albergel, C., Arduini, G., Beljaars, A., Bidlot, J., ... Zeng, X., 2018: Satellite and in situ observations for advancing global Earth surface modelling: A review. *Remote Sens.*, 10(12), 2038. https://doi.org/10.3390/rs10122038

Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., ... Tiede, D., 2014: Geographic object-based image analysis – towards a new paradigm. *ISPRS J. Photogramm. Remote Sens.*, 87, 180–191. https://doi.org/10.1016/j.isprsjprs.2013.09.014

Böttcher, H., Urrutia, C., Benndorf, A., Martius, C., Atmadja, S.S., Boissière, M., ... Kalman, R., 2025: Transparent monitoring in practice: A guide to effective monitoring in the land sector. Book II – Case studies of transparent monitoring approaches.

D'Odorico, P., Gonsamo, A., Damm, A., Schaepman, M.E., 2013: Experimental evaluation of Sentinel-2 spectral response functions for NDVI time-series continuity. *IEEE Trans. Geosci. Remote Sens.*, 51(3), 1336–1348. https://doi.org/10.1109/TGRS.2012.2224118

European Environment Agency, 2018: CORINE Land Cover 2018. Retrieved from https://land.copernicus.eu/pan-european/corine-land-cover

Foody, G.M., 2002: Status of land cover classification accuracy assessment. *Remote Sens. Environ.*, 80(1), 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4

Gobron, N., Pinty, B., Verstraete, M.M., Widlowski, J.L., 2002: Advanced vegetation indices optimized for up-coming sensors: Design, performance, and applications. *IEEE Trans. Geosci. Remote Sens.*, 38(6), 2489–2505. https://doi.org/10.1109/36.885202

Grace, J., 2004: Role of forest biomes in the global carbon balance. In: *The Carbon Balance of Forest Biomes*, Taylor & Francis, 19–45.

Hasan, M.F., Sojib, M.R., 2025: Climate change resilience. *Intelligent Solutions to Evaluate Climate Change Impacts*, 347.

He, K., Zhang, X., Ren, S., Sun, J., 2016: Deep residual learning for image recognition. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Helber, P., Bischke, B., Dengel, A., Borth, D., 2019: EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 12(7), 2217–2226. https://doi.org/10.1109/JSTARS.2019.2918242

Ienco, D., Gaetano, R., Dupaquier, C., Maurel, P., 2017: Land cover classification via multitemporal spatial data by deep recurrent neural networks. *IEEE Geosci. Remote Sens. Lett.*, 14(10), 1685–1689. https://doi.org/10.1109/LGRS.2017.2734770

Immitzer, M., Vuolo, F., Atzberger, C., 2016: First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sens.*, 8(3), 166. https://doi.org/10.3390/rs8030166

Kaplan, G., Avdan, U., 2017: Mapping and monitoring wetlands using Sentinel-2 satellite imagery. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-2/W4, 271–277. https://doi.org/10.5194/isprs-annals-IV-2-W4-271-2017

Kelly, A.C., 2017: Improving REDD+ (Reducing Emissions from Deforestation and Forest Degradation) programs. Doctoral dissertation.

Köhl, M., et al., 2015: Forest Resources Assessment 2015: How are the world's forests changing? Food and Agriculture Organization of the United Nations. Retrieved from http://www.fao.org/forest-resources-assessment/en/

Loschilov, I., Hutter, F., 2016: SGDR: Stochastic gradient descent with warm restarts. *arXiv preprint*, arXiv:1608.03983. https://doi.org/10.48550/arXiv.1608.03983

Lu, D., Weng, Q., 2007: A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.*, 28(5), 823–870. https://doi.org/10.1080/01431160600746456

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019: Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.*, 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

Macarringue, L.S., Bolfe, É.L., Pereira, P.R.M., 2022: Developments in land use and land cover classification techniques in remote sensing: A review. *J. Geogr. Inf. Syst.*, 14(1), 1–28. https://doi.org/10.4236/jgis.2022.141001

Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., ... Wu, H., 2017: Mixed precision training. *arXiv preprint*, arXiv:1710.03740. https://doi.org/10.48550/arXiv.1710.03740

Simonyan, K., Zisserman, A., 2014: Very deep convolutional networks for large-scale image recognition. *arXiv preprint*, arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

Tan, M., Le, Q., 2019: EfficientNet: Rethinking model scaling for convolutional neural networks. In: *Proc. Int. Conf. Mach. Learn. (ICML)*, PMLR, 6105–6114.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F., 2017: Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.*, 5(4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307