SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V12I5P105

Volume 12 Issue 5, 45-57, May 2025
© 2025 Seventh Sense Research Group®

Review Article

A Review of Deep Learning Approach for Analyzing Remote Sensing Spectral Data in Species

Nita Nimbarte[1], Vandana Katna[1], Sanjay Balamwar[2]

[1]*Department of Electronics & Telecommunication Engineering, Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, India.*
[2]*Maharashtra Remote Sensing and Applications Centre (MRSAC), VNIT Campus, Nagpur, India.*

[1]*Corresponding Author : nitangp@gmail.com*

Received: 28 February 2025 Revised: 05 April 2025 Accepted: 19 April 2025 Published: 31 May 2025

Abstract - *Remote sensing technologies have become crucial for forest management, providing large-scale data through satellite and aerial imagery. Automated semantic segmentation of trees enables efficient monitoring, although the task remains challenging due to varying tree spectral signatures, the limited availability of labeled datasets, and geometric distortions. In the domain of precision agriculture, major research efforts have focused on monitoring agricultural fields, classifying land use, and optimizing crop yields. In terms of accuracy and reliability, deep learning algorithms now perform noticeably better than conventional machine learning techniques for remote sensing image analysis. Recent advances in image segmentation, a key field in computer vision, have enabled more accurate identification and categorization of objects in remote sensing images. While many studies rely on frontal or asymmetrical image views, this review focuses on deep-learning approaches using top-down datasets for species and land cover segmentation. Models such as U-Net, SiU-Net, and DeepLabV3+ demonstrate notable performance improvements, achieving mean average precisions of 0.921, 0.970, and 0.976, respectively. Compared to earlier conventional approaches, these models show a significant leap in accuracy, particularly in handling fine-grained details and large-scale environmental variations. Furthermore, independent validation using tree species proportion maps highlights the practical reliability of these models in estimating species presence, absence, and distribution, thereby reinforcing their importance in advancing remote sensing-based ecological and agricultural monitoring.*

Keywords - *Remote Sensing, Multispectral Data, Hyperspectral Data, Deep Learning, Image Classification.*

1. Introduction

The forest ecosystems provide vital services, necessitating sustainable management policies to maintain their ecological and economic roles. Effective strategies require understanding regional and national forest dynamics, particularly under climate change [1]. Remote sensing, alongside field inventories, efficiently gathers forest data. Machine learning and deep learning have shown promise in satellite-based environmental monitoring [2]. Mapping Land Use and Land Cover (LULC) is crucial for tracking urban sprawl, negatively affecting agriculture, water infiltration, and open spaces. Current LULC mapping methods use deep learning, machine learning and satellite imagery techniques [3]. Forests are critical for climate stability, making their accurate assessment a global priority. Remote sensing surpasses traditional methods like aerial surveys in monitoring forest changes. Previously, manual analysis dominated, but recent advances in artificial intelligence introduced Deep Learning (DL) and satellite technologies [4]. Many countries, particularly in Europe, rely on land cover classification and management to boost agricultural yields. Land cover maps provide spatial data on features like forests and croplands, while dynamic maps capture temporal changes [5]. Land use maps detail human activity affecting land cover.

Improved spatial resolution enables more detailed satellite image analysis, shifting from object-based to pixel-level semantic segmentation. Sentinel-2's publicly available multispectral imagery has advanced land-cover research [8]. Over the last two decades, deep learning has enhanced computer vision tasks like preprocessing, segmentation and scene understanding. However, DL techniques depend on extensive, high-quality datasets. RGB imaging struggles in low light or fog, necessitating alternative techniques [9]. Image segmentation is vital for visual-based applications, including remote sensing, autonomous driving, medical imaging, and augmented reality [11]. Machine Learning (ML), a key artificial intelligence subset, underpins many remote sensing analyses. Numerous ML algorithms have been developed for various applications in recent years [13]. The study reviews previous research on remote sensing, satellite imagery, and learning models, detailing the classification

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

methods and modal architectures employed. It analyses model performance based on accuracy results, highlighting the best-performing models. Additionally, it discusses the advantages, limitations, and notable strengths of the models evaluated. This revision explicitly highlights the research gap in terms of current limitations in forest monitoring and LULC mapping methods. It also introduces the problem of achieving accurate assessments in the context of low-resolution data, climate change, and the need for improved deep learning and satellite-based techniques.

2. Methodology

2.1. Land Cover Segmentation

The methodology is separated into two sections: the characteristic designs of the suggested SiU-Net and traditional models (DeepLabV3+ and U-Net) and their performance evaluation with test data. The deep learning models section focuses on network architecture, especially the input nodes and encoders for SiU-Net. Comparing model performance qualitatively and quantitatively, validating SiU-Net's superiority and reviewing each model's architecture and class-specific features.

DeepLabV3+ and U-Net, both encoder-decoder models, are widely used for semantic segmentation due to their ability to restore localization and boundary details. SiU-Net, built on U-Net, incorporates ReLUs, Batch Normalization (BN), and 3 × 3 convolutions [1]. The DeepLab series tackles multiscale issues using Atrous Spatial Pyramid Pooling (ASPP). DeepLabV3+ enhances accuracy and speed by expanding the receptive field without increasing parameters or reducing spatial resolution, and it refines low-level features for better boundary detection [7].

Multispectral imagery offers a broader spectral range, improving segmentation over RGB, especially for water bodies. However, using all 20 channels provided limited gains in the Kaggle challenge, where upsampling lower-resolution bands risks data loss. Advanced models like SharpMask, U-Net, and RefineNet embed hierarchical features into decoders for improved segmentation, while DeepLabV3+ leverages spatial pyramid pooling for multiscale feature extraction.

Multispectral images, captured using specialized instruments or filters, span RGB, near-infrared (750–900 nm), and thermal infrared (10410–12510 nm) ranges. Given the time-intensive nature of data collection and the need for large datasets for Deep Convolutional Neural Networks (DCNN), data augmentation was used to increase data volume, enhance regularization, and reduce overfitting. Techniques like Translation, Rotation, Flip, Crop, Scale and Gaussian noise added diversity. Annotation, essential for DCNNs, was performed using the bounding box method with free, open-source tools. The annotations were stored in darknet format for YOLO v3 compatibility [9].

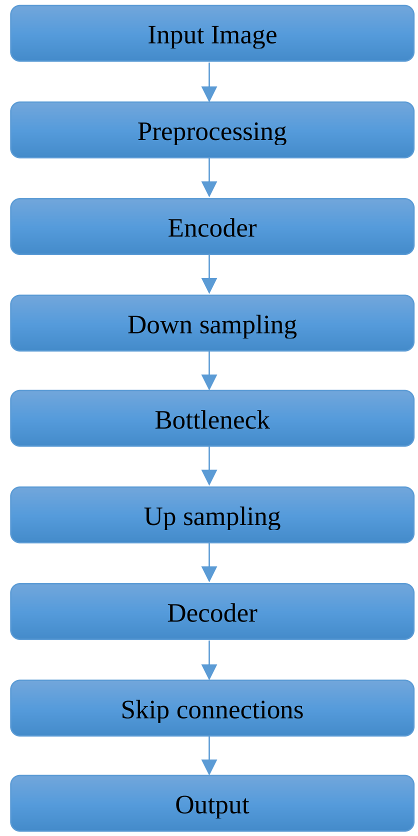

Fig. 1 Block diagram of U-Net method. [1, 7-11]

Figure 1 depicts a standard architecture for semantic segmentation models designed to assign a class label to every pixel in the input image. The model employs an encoder-decoder structure: the encoder extracts features through convolutional layers and down-sampling, capturing high-level semantic information, while the decoder up-samples these features to generate a pixel-wise segmentation map. A crucial component is the bottleneck, where the most abstract representation of the image is processed. The final output has finer and more accurate segmentations due to skipping connections between the relevant encoder and decoder layers, which aid in recovering fine-grained features lost during down. Preprocessing steps are often applied to the input image before feeding it to the network. Table 1 demonstrates the comparative analysis of three semantic segmentation architectures: DeepLabv3+, U-Net, and SIU-Net. It highlights their key architectural features, techniques used, context and resolution handling, typical use cases, and performance characteristics. DeepLabv3+ excels in complex scenes with its ASPP module; U-Net is efficient for precise segmentation, especially in medical imaging, while SIU-Net uses attention mechanisms for fine-grained detail segmentation. Sampling strategies in machine learning select relevant training data, enhancing speed and accuracy. In remote sensing, they identify representative sites over large areas. Low-resolution satellite data like MODIS (500 m per pixel) supports global mapping, while Landsat provides detailed land cover analysis.

Table 1. Key differences summary of DeepLabV3+, U-Net, and SiU-Net. [1, 7-11]

Feature	DeepLabV3+	U-Net	SiU-Net
Architecture	Encoder-Decoder with ASPP and refinement.	Encoder-decoder with skip connections.	U-Net variant with spatial attention.
Key Technique	Atrous convolutions (dilated) + ASPP	Skip connections, upsampling (transposed convolution).	Attention mechanisms for region focus.
Context Handling	Multi-scale context via ASPP	Strong local context retention via skips.	Focuses on fine details with attention.
Resolution Handling	Multi-scale context with global receptive fields.	Maintains high spatial resolution with skips.	Improves segmentation of fine details through attention.
Use Cases	General semantic segmentation (e.g: urban, large-scale)	Medical image segmentation, fine boundaries.	Single image segmentation, small object focus.
Performance	Strong on large datasets, and complex scenes.	Efficient for small datasets, precise segmentation	Optimized for accuracy on fine-grained tasks.

Sentinel-2 acquires multispectral imagery across 13 bands, spanning the visible to shortwave infrared spectrum, with spatial resolutions resampled to 10 meters. Traditional machine learning methods, like K-Means clustering, are effective for land cover classification. Pseudo-labeling assigns probabilities to pixels when labeled data is scarce, setting zero for misclassified pixels.

U-Net [1,7,8], FPN, and DeepLab with a ResNet50 backbone and 128 × 128 input size are commonly used for segmentation. Implemented using Pytorch, these architectures show strong performance in remote sensing. Precision, recall, and F1-score assess model quality, while IoU (Jaccard Index) measures segmentation accuracy by dividing the intersection of predicted and ground truth masks by their union [10]. Semantic segmentation assigns class labels to each pixel, combining object detection with segmentation. FCN is foundational for models like U-Net, which is known for its encoder-decoder structure and skip connections.

DeepLabV3 enhances segmentation with atrous convolution and Spatial Pyramid Pooling (SPP) for multiscale feature handling, modifying ResNet-101 to retain high-resolution features. It splits feature maps into spatial bins to manage various input sizes. U-Net also follows an encoder-decoder structure but uses a symmetric design to avoid direct connections between upsampling and downsampling paths. DeepLabV3 combines parallel and cascaded atrous convolution modules, using color variables for top-view human segmentation. These models reconstruct outputs via upsampling techniques, integrating atrous separable convolutions to capture context without extra parameters [11].

Figure 2 describes the block diagram of a semantic segmentation method incorporating an Atrous Spatial Pyramid Pooling (ASPP) module. The process begins with an input image, which is preprocessed before being fed into a backbone network (e.g., ResNet, MobileNet). The backbone extracts hierarchical feature maps, which are then passed through the ASPP module. ASPP utilizes several parallel atrous convolutions with different dilation rates to extract contextual information at multiple spatial scales. ASPP output is then fed into a decoder to generate the final output, a pixel-wise segmentation map highlighting different objects or regions in the image.

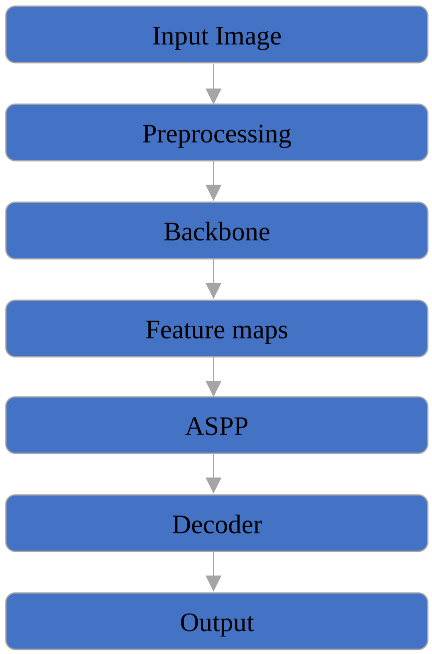

Fig. 2 Block Diagram of DeepLabV3+ method. [1, 7-11]

2.2. Spectral-Spatial Deep Learning

The decoder uses transposed convolutions (upsampling) with skip connections that link encoder and decoder features at each level, maintaining identical input-output dimensions. UNet++ enhances segmentation accuracy over UNet by introducing deep supervision and refined skip paths, which reduce the semantic gap and support multi-level concatenation. It also uses weight regularization to mitigate overfitting and boost generalization. Evaluated loss functions include Kullback-Leibler divergence, pseudo-Huber (PH), mean absolute error, mean squared error and focal loss. Data processing and evaluation utilized the raster, sf, keras (TensorFlow), GDAL/OGR, and Orfeo ToolBox libraries [2].

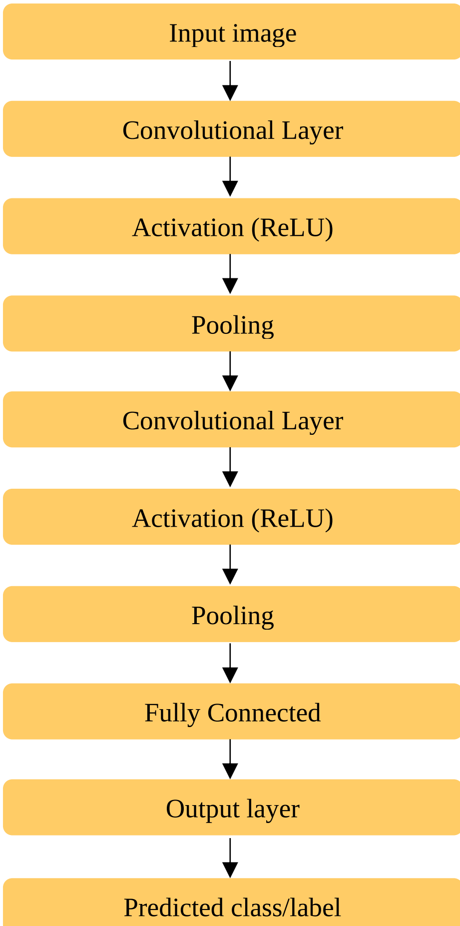

Fig. 3 CNN architecture diagram. [2, 12-16]

Deep Learning (DL) models are composed of M layers, each containing N nodes, which enable the step-by-step representation of data. A fully connected network, also known as a Multilayer Perceptron (MLP), connects all nodes across layers using weighted connections. Using multitemporal satellite imagery, MLP classifies poplar plantations using data from the 2017 polygon. TensorFlow optimized the MLP and trained on 2000 polygons (1000 each for poplar and non-poplar) through trial and error. A 17-layer MLP with varied nodes and activation functions achieved better accuracy. The first layer applied a threshold to set negatives to zero; the second used Tanh activation. Dropout layers prevented overfitting by disabling random nodes [12].

Figure 3 depicts the architecture of a basic CNN model for image classification. The network begins with an input image, which is processed through a series of layers. First, convolutional layers apply learnable filters to extract spatial hierarchies of features. The network can learn intricate patterns due to the non-linearity introduced by ReLU activation functions. In order to reduce computational complexity and provide some invariance to slight spatial variations in the input, pooling layers downsample the feature maps. These convolutional and pooling layers are repeated to form a feature extraction stage. Finally, the extracted features are passed through fully connected layers that perform high-level reasoning and ultimately output a predicted class/label for the input image. DL frameworks evolve rapidly, offering efficient training with automated gradients for CPU/GPU and prebuilt neural network classes. CNN-based studies in LULC classification benefit from rich datasets. Remote Sensing (RS) CNN research covers image segmentation and object detection. Usage analysis shows 60% of segmentation studies use multispectral satellite images, followed by aerial (18%), multi-data (8%), UAV (7%), RADAR (4%), LiDAR (2%), and panchromatic (1%) images. Open datasets significantly influence land cover research.

Other studies utilized specialized space-borne datasets for large-scale research, with Google Earth being the most common source, followed by Worldview-4, Gaofen1-2, and Quickbird-2. Region-based CNN (R-CNN) architectures, such as fast R-CNN, faster R-CNN, and mask R-CNN, were the most prominent in the reviewed designs. Land cover mapping comprised 39% of the applications, followed by agriculture (15%), urban areas (11%), wetlands (12%), forests (10%), disasters (3%), and soil (2%).

Approximately 8% of the case studies focused on geology, water mapping, benthic habitats, rock types, and mining classification. Among backbone models, VGG variations were the most popular (34%), followed by ResNet (30%), with less use of Inception, SegNet, LeNet, and GoogleNet. CNNs also supported tasks like image registration, change detection, data fusion, and super-resolution. Their effectiveness in panchromatic/multispectral data fusion stems from their ability to model complex data relationships. In recent years, DL methods have performed well in satellite image change recognition applications [13].

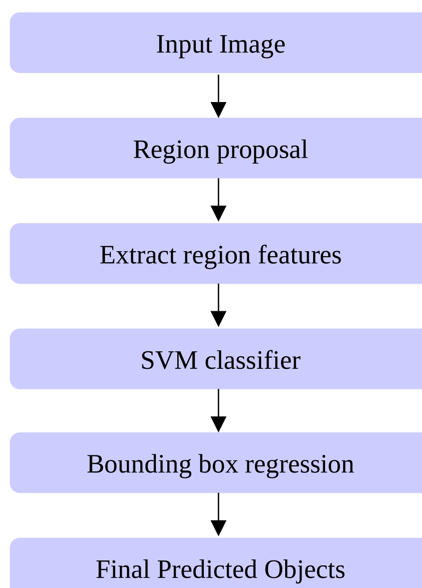

Fig. 4 SVM Architecture Diagram. [2, 12-16]

Figure 4 describes a standard object detection pipeline utilizing a Support Vector Machine (SVM) classifier. The process begins with an input image and generates region proposals, which are potential bounding boxes that may contain objects. Features are then extracted from these regions, typically using methods such as Histogram of Oriented Gradients (HOG) or deep learning-based feature extractors. Finally, an SVM classifier evaluates each region to determine whether it contains an object of interest. Finally, bounding box regression is employed to refine the location and size of the predicted bounding boxes, resulting in the final predicted objects with accurate locations and class labels.

Pan-sharpening combines a high-resolution Panchromatic (PAN) image with a Multispectral (MS) image to produce a composite with improved spectral and spatial details. MS images contain multiple spectral bands, while PAN images have a higher spatial resolution but only one band. Pan-sharpening techniques include spectral, spatial, and spatial-spatial methods, enhancing data integration from sensors and platforms. Deep Learning (DL), particularly neural networks, has been applied in remote sensing for object detection, classification, and image fusion. Sentinel-2 provides 13 spectral bands at varying resolutions, enabling fine-scale analyses and detailed change detection despite spatial resolution limits [14].

The presented methodology involves preprocessing multi-temporal Sentinel-1 and Sentinel-2 data, extracting reference data from the National Forest Inventory (NFI), and developing machine learning models through parameter tuning and validation. Image composites for winter (post-deciduous senescence) and summer (post-foliation) reflect seasonal landscape changes. Google Earth Engine (GEE) facilitated satellite imagery preprocessing, including NDVI computation and noise removal using median composites. Selected bands (visible, NIR, and SWIR) supported a hierarchical classification process for tree species, forest type and forest cover. A Random Forest (RF) classifier handled supervised pixel categorization, mitigating overfitting with its ensemble decision tree approach [15]. Hyperspectral data identified tree species, emphasizing European aspen.

The process involved individual tree delineation, ground-reference reconciliation, model fitting, and unlabeled tree classification. LiDAR-derived CHM filtered tree height and shadows, applying the Dalponte and Coomes algorithm for crown delineation. CNN models used square patches around treetops for classification, testing patch sizes from 4 to 10 m. Larger patches included multiple trees, enhancing classification potential. CNNs handle hyperspectral data by extracting features from pixel spectra (1D-CNN) or spatial dimensions (2D-CNN), though 2D-CNNs increase parameters without fully leveraging spectral data. 3D-CNNs, combining spatial and spectral information, demonstrated superior results by generating feature cubes. DL models were compared with ANN, RF, SVM, and GBM for remote sensing tasks. Implementations used NVIDIA V100 GPGPU with PyTorch, fastai2, and Light GBM, highlighting CNN applications in RGB imagery interpretability [16].

2.3. Land Use and Land Cover

The Mini France dataset provides VHR aerial images and labels, serving as the first benchmark for semi-supervised learning in Land Use and Land Cover (LULC) classification. Sentinel-2 data includes RGB and NIR bands at 10 m/pixel, while other bands range from 20 to 60 m/pixel. High label quality presents challenges with temporal changes and invisible classes. CNNs have proven effective for LULC mapping but require substantial training data. Random Forest (RF) is a popular alternative, offering fast training, resistance to overfitting, and straightforward feature importance analysis. Geographic Object-Based Image Analysis (GEOBIA) categorizes image pixels using object segmentation, leveraging texture and spectral characteristics through Simple Non-Iterative Clustering (SNIC) [3].

For SAR data preprocessing, SNAP toolbox methods and the GLCM module compute texture features from VV and VH polarizations, aiding the classification of land cover groups such as forests, cropland, and urban areas. Object-based classification using eCognition® v.9.01 follows a three-step process: segmentation, object hierarchy creation, and classification. Parameters like scale and compactness help identify homogenous land patches, producing realistic classifications comparable to natural-colour Sentinel-2 composites [17].

GEOBIA emphasizes geo-centric, multi-source, and context-aware analysis. It uses segmentation to divide data into homogeneous regions, serving as a step toward extracting meaningful objects. Traditional segmentation includes pixel-, edge-, and region-based methods. GEOBIA distinguishes between scenes (real-world objects) and images (sensor representations). Unlike the geo-relational model, which separates spatial and attribute data, the object-oriented data model treats real-world entities as interconnected objects. Optimization in GEOBIA aims to create image objects that align with predefined ontologies, classifying segments by spatial, spectral, and topological properties rather than individual pixels [18].

Sentinel-1 imagery was sourced from the Alaska Satellite Facility (ASF), while Sentinel-2 data was accessed and preprocessed using the Google Earth Engine (GEE) Python API. Preprocessing of Sentinel-1 data was conducted using SNAP version 8.0.3, an open-source tool developed by the European Space Agency (ESA). Classification tasks were performed using the Scikit-learn Python library. The SAR dataset comprised time-series Sentinel-1A/B Ground Range Detected (GRD) images with VV and VH polarizations, acquired in Interferometric Wide (IW) mode.

Bilinear interpolation was used to resample both the final image and the DEM. The RF classifier, implemented using Scikit-learn's RF function, was applied for pixel-based forest cover classification. RF, a decision-tree-based machine learning method, is widely used for land cover mapping and forest classification. To enhance ecosystem monitoring and vegetation analysis, combined S1 and S2 time series were evaluated for classifying forest tree species, aiding short- and long-term disturbance impact assessments [19].

Table 2 summarises the workflow for land cover classification using remote sensing data. It starts with data acquisition and preprocessing, followed by selecting training data for supervised methods. Feature extraction and classification are then performed, with post-processing refining the results. Finally, accuracy is assessed, and the classified data is interpreted for applications such as land use planning.

Table 2. LULC Classification workflow [3, 17-21]

Steps	Description
Data Acquisition	Collect remote sensing data (satellite imagery, aerial photos, LiDAR).
Pre-processing	Process raw data (radiometric correction, geometric correction, cloud removal).
Training Data Selection (for supervised classification)	Manually select sample regions that represent different land use/cover classes.
Feature Extraction	Extract relevant features (spectral bands, vegetation indices like NDVI).
Classification	Apply classification algorithm (supervised, unsupervised, object-based, or deep learning).
Post-Processing	Refine results with techniques like majority filtering, smoothing, or object merging.
Accuracy Assessment	Assess classification accuracy using metrics like confusion matrix, Kappa coefficient, etc.
Interpretation	Analyze results for decision-making (e.g; land use planning, environmental monitoring).

An aerial LiDAR scanner captured the orthophoto, which required geometric calibration using 34 Ground Control Points (GCPs) from identifiable locations like corners and power lines. ArcGIS 10.5 was used for the geometric adjustment. The goal of classification models is to label each pixel based on training samples and ground truth data, typically using spectral information. Object-Based Image Analysis (OBIA) can also enhance classification by dividing the image into homogeneous groups, incorporating texture, shape, and spatial features. However, both pixel-based and OBIA methods face challenges like segmentation optimization and speckle noise. Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have been introduced to overcome these. CNNs, which simulate human vision through shared weights and local receptive fields, were used to classify ultra-high-resolution aerial orthophotos, with performance evaluated using overall accuracy, Kappa coefficient, and per-class accuracy. The CNN model used a convolutional layer, two dense classifiers, batch normalization, and max pooling, with training via stochastic gradient descent and backpropagation [20].

Sentinel-2's Multispectral Instrument (MSI) includes bands with spatial resolutions of 10 m, 20 m, and 60 m, with 10 m being the highest available. It also features three red-edge bands for detecting vegetation reflectance in the near-infrared region. Spectral indices like NDVI, MNDWI, and NDBI were used to enhance the categorization of vegetation, water, and artificial surfaces, respectively. The SVM method finds the optimal decision boundary by identifying training data patterns and applying them to evaluation data. RF is an ensemble learning algorithm that uses bootstrap aggregation for improved predictions. Gradient Boosting Machine (GBM) optimizes weak decision trees to generalize models. A multilayered Deep Neural Network (DNN) with error backpropagation and hyper-parameter optimization was employed for classification, with the Softmax function normalizing each class's output to generate a probability distribution [21].

2.4. Semantic Segmentation

In precision agriculture, the goal is to distinguish healthy trees, aided by the red edge (0.71–0.75μm). Sick trees are identified by detecting crown yellowness using the yellow band (0.59–0.63μm). The NIR2 band (0.86–1.04μm), less influenced by atmospheric conditions, provides better vegetation data. Spatial resolution impacts segmentation performance, particularly the Ground Sampling Distance (GSD) and the sensor's ground Field-of-View (FOV). The WorldView-3 satellite captures images in various resolutions, including panchromatic, RGB, multispectral, and SWIR. Deep learning-based semantic segmentation models like U-Net, SegNet, and DeepLabv3+ are widely used. U-Net, with four symmetric layers, utilizes a ReLU activation function and batch normalization. DLinkNet, a model with dilated convolution layers, is designed for near-zero zenith angle images, such as those used in road segmentation. DeepLabv3+ consolidates multiscale contextual information using atrous spatial pyramid pooling and a lighter decoder. RF is introduced for performance evaluation, comparing deep learning and traditional machine learning approaches [4].

Deep learning and UAVs are crucial in automated forest monitoring. Images were captured at 100 meters and processed with Agisoft Metashape, and GPS-located crowns were identified using orthomosaics. Two indices, normalized difference and normalized bands, were computed for crown pixels. The Mask R-CNN pipeline was used to classify image tiles into 600 x 600 pixel patches [22]. The Hybrid-CNN model extracts both spatial and spectral data with components like 2D and 3D CNN modules and a Softmax classifier. Band selection techniques (RGB, NIR, RGB+NIR, UBS-9) reduced dimensionality to address spectral redundancy. Classification performance was evaluated using metrics such as overall accuracy, precision, and kappa coefficient [23]. The RF model utilizes a stepwise RFE process to eliminate irrelevant variables, improving prediction accuracy. Bands selected for RapidEye, Sentinel-2, PlanetScope, and Landsat 8 were optimized for various tasks. RF regression models predicted tree species diversity based on field data, with statistical metrics like RMSE and R2 used to evaluate model performance. The RF algorithm's variable relevance feature helped assess predictors for improved performance [24].

Surface characteristics exhibit unique spectral reflectance signals, similar to how fingerprints are used for human identification due to their consistency over time. Different surface objects absorb and reflect electromagnetic radiation based on their physical and chemical properties. Spectral reflectance measures the energy a surface reflects at specific wavelengths. The ASD Fieldspec 4 utilizes three spectral detectors that cover the wavelengths of 350-1000nm, 1001-1800 nm, and 1801-2500 nm. However, spectral discontinuities may arise at the boundaries of these ranges due to target inhomogeneity and sensor warm-up issues. A radiometric inter-channel jump correction method helps to mitigate these discrepancies. High spectral resolution sensors provide superior feature identification compared to multispectral sensors. To assess the potential performance of these sensors in vegetation classification and unmixing, synthetic remote sensing images were paired with field-surveyed data.

Table 3 describes reviews of various image processing or analysis approaches, linking their architectural choices with feature selection methods and relevant vegetative indices. It covers deep learning-based semantic segmentation using architectures like U-Net and DeepLabv3+, along with traditional Convolutional Neural Networks (CNNs). Random Forest Regression and spectral unmixing are also included, showing their preferred data sources and indices. The table highlights the diverse tools and techniques available for analyzing image data, particularly in the context of vegetation studies.

Real hyperspectral sensors typically have wider bandwidths (5-10 nm), but the synthetic hyperspectral image created for this study had 2101 spectral bands (400-2500 nm) with a 1 nm spectral resolution. To make the hyperspectral images more realistic, a Gaussian spectral response function was applied to aggregate the 1 nm spectra into 10 nm bands. The classification was performed using the Spectral Angle Mapper (SAM), a physical-based classifier that matches pixels to reference spectra through an n-dimensional Spectral Angle (SA).

This method is immune to illumination or albedo effects and links directly to end member spectra from the spectral library. Each pixel was assigned to the class with the smallest SA [25]. The spectrometer used in this study had a wavelength range of 350–2500 nm, with spectral resolutions of 3–1.4 nm for 350–1000 nm and 6–2 nm for 1000–2500 nm. Spectra were collected under dry, sunny, and windless conditions. These spectra reflect various biophysical characteristics of vegetation, including chlorophyll content, canopy structure, nitrogen content and stress-related pigments. Four classification models, RF, SVM with radial basis function kernel, Back Propagation Neural Network (BPNN), and Regularized Logistic Regression (RLR), were implemented using the CRAN R package caret [26].

Table 3. Description of classification, feature selection, and vegetative indices [4, 22-26]

Architecture	Feature Selection	Vegetative Indices
DL-based semantic segmentation	DSTL image and RIT image.	NDVI, ARVI, SAVI, RGB, NIR
CCN architecture	-	RGB, NIR, RGB+NIR, USB-9
Random Forest Regression Modelling	Landsat8 image, Planet Scope image, Rapid Eye, Sentinel2 image	Shannon index, Simpson index and Species richness
Classification and spectral unmixing of synthetic images using spectral library	Hyperspectral and Multispectral	VNIR, SWIR and VNIR-SWIR. IKONOS, Landsat-8 and WorldView-3

2.5. Image Segment using Deep Learning Model

The ResNet-50 model was the second used, incorporating various weight initialization techniques such as random weights and pre-trained ResNet on the ImageNet dataset. This transfer learning approach aids in applying prior knowledge to more challenging tasks due to limited training data. The satellite segmentation model utilized the DeepGlobe dataset, enabling the use of learned ResNet weights, with adjustments made to the final layers. UNet, known for its scalability in semantic segmentation, consists of two paths: a contracting path for context and an expanding path for precise localization. The contracting path mimics ResNet's architecture with skip connections, and the expanding path improves feature map resolution via transposed convolution. UNet generates the resulting pixel-wise mask. ResNet-50 was used as the encoder with 48 convolution layers and one MaxPool layer. ResNet helps avoid degradation and vanishing gradient issues due to its skip connections in ResBlocks, allowing deeper networks without performance loss [5]. The model integrates CNN and FPCRF components, optimizing segmentation through coadaptation. The CNN outputs segmentation probabilities and feature embedding for pairwise potential computation, promoting similar labeling for pixels with similar characteristics. FPCRF enhances CNN output by modeling spatial correlations between unary potential and feature embedding, ultimately outputting the marginal distribution for each pixel's class label [27].

Based on BDL, the model consists of a segmentation component (F) and a translation component (F). The segmentation part labels the input images and includes a Domain Discriminator (DM) that distinguishes between target and source datasets. The translation network, a Cycle GAN with two ResNet generators and discriminators, translates images between datasets while maintaining consistency through cycle loss. The segmentation network, based on DeepLabV2 and ResNet101, utilizes ImageNet-pre-trained weights. The domain discriminator reduces domain shift, encouraging consistent labels across datasets. Cloud-Net, used for cloud-covered data, features a U-net-like architecture with six convolution blocks and five deconvolution blocks. It was trained with Landsat rasters in four spectral layers (red, green, blue, NIR) [28].

WorldView-3, a commercial satellite, provides high-resolution imagery: 7.5 m for short-wave infrared, 1.24 m multispectral, and 31 cm panchromatic resolution. Panchromatic sharpening combines lower-resolution M-band images with higher-resolution panchromatic images, producing an M-band image with panchromatic resolution when rasters overlap. The U-Net architecture is composed of a contracting path and an expanding path. The contracting path follows a typical convolutional neural network structure, incorporating batch normalization to accelerate convergence during training. Instead of the commonly used ReLU, the Exponential Linear Unit (ELU) is employed as the activation function, enhancing learning and increasing robustness to noise. With each downsampling step, the number of feature channels is doubled. The expanding path performs upsampling, applies a convolution with fewer feature channels, concatenates the result with the corresponding feature map from the contracting path, and follows up with batch normalization and ELU activation [29].

Table 4 describes various deep learning architectures and their applications. It details using U-Net with ResNet-50 for segmentation and classification, combining CNNs with Fully connected Conditional Random Fields (FPCRF) for footprint generation and distribution analysis. The table also mentions a Bidirectional Long short-term memory (BDL) model for translation and domain discrimination, Fully Convolutional Networks (FCNs) for image classification, and basic CNNs for unspecified tasks, potentially including rapid MFCC algorithm calculation. Each row links a specific architecture with its corresponding task and relevant techniques. In speech recognition, two primary components, phonograms and spectrograms, work together to process speech. A key task is identifying the start and end of sentences in noisy environments or isolating speech sections from the signal. Zero-crossing, where the function’s sign changes, helps detect useful signals amidst noise, which can degrade recognition performance. Digital filters (line and beginning filters) are used to minimise noise. Speech signal segmentation into frames, known as “segment-stations”, is common. The rapid Fourier transform converts the temporal speech signal into a spectral frequency, aiding in separating speech from noise and handling acoustic variations. The MFCC algorithm, using spectral and phonogram data, applies K-Nearest Neighbors (KNN) for extracting valuable information [30]. They adapted ILSVRC classifiers for segmentation into Fully Convolutional Networks (FCNs), enhancing them with pixel-wise loss and in-network upsampling. Training used the PASCAL VOC 2011 segmentation challenge, with a multinomial logistic loss per pixel and mean pixel Intersection over Union (IoU) for validation. The FCN architecture improves spatial precision by incorporating the feature hierarchy, and a dual-headed variant, trained for both semantic and geometric predictions, outperforms two independent models in both tasks, with learning and inference speeds comparable to each individual model [31].

Table 4. Description of DL algorithm from reference papers [5, 27-31]

Architecture	Description
UNet model and ResNet-50 model	UNet model used for segmentation task, ResNet-50 model for both classification and encoder for the modified UNet model.
CNN and FPCRF model	CNN model for embedding, for building foot print generation is FCN, graphical representation of distribution are Bayesian networks and Markov Random Field (MRF), FPCRF for marginal distribution.
BDL model	BDL model is divided into two parts are translation part F and Domain Discrimination (Dm), cloud masking network.
Fully convolution networks	ILSVRC classifier into FCNs. Multispectral U-Net architecture, reflectance indices.
CNN	CNN in 1DCNN and 2DCNN, for rapid calculation MFCC algorithm is used.

2.6. Multispectral and Hyperspectral Imagery

Sentinel-2 (S-2) and WorldView-2 (WV-2) multispectral sensors, offering spatial resolutions of 10-60 meters and 0.46-1.84 meters, respectively, and capturing data across 13(S-2) and 8(WV-2) spectral bands, provide satellite imagery that is coordinated in acquisition timing with hyperspectral airborne and RPAS over-flights. In-situ field data was collected during these missions for training and validation. Reflectance measurements (350-2500 nm) were taken using the ASD FieldSpec-3 spectro radiometer over homogeneous areas, with accurate sampling to capture species variability. A pre-processing step includes atmospheric correction, with the RPAS radiometer measuring irradiance (400–1000 nm) for the Pika–L scene. Principal Component Analysis (PCA) and Minimum Noise Fraction (MNF) were applied to hyperspectral images to extract key information. Maximum Likelihood (ML) and SVM classifiers outperformed others like Spectral Angle Mapper (SAM) in seabed mapping. SVM is recognized for its robustness and high accuracy relative to methods such as deep convolutional neural networks and random forests. Mapping accuracy was assessed using the Kappa coefficient and a confusion matrix. The Gram-Schmidt algorithm was employed to enhance spatial detail, while atmospheric correction was performed using FLAASH. The Minimum Noise Fraction (MNF) transformation was also used to reduce data redundancy [6].

The method consists of two stages: 1) Super-resolution of MS images and 2) Fusion with PAN images. LRMS images are enhanced through a fast iterative super-resolution technique, followed by hierarchical fusion to preserve spectral information and spatial details. The YUV components reduce spectral distortion by matching histograms between the PAN and Y components of MS super-resolution images. The final high-resolution Multispectral (MS) image is generated by merging the modified brightness component with the U and V components through an inverse YUV transformation [32]. A feature extraction approach based on Discrete Modal Decomposition (DMD) was implemented, incorporating both full-scale and filtered DMD techniques. The process involves training the algorithm on labeled data and then predicting labels for new data. The classification performance is measured by the accuracy of predictions [33].

PRISMA products are categorized based on utility. Level 0 includes satellite data, and Level 1 consists of radiance images and hypercubes calibrated geometrically and radiometrically. Level 2 has sub-levels, with L2A providing cloud masking and land cover mapping and L2B offering atmospheric elements like aerosols and water vapor. The data is available in HDF5 format, with geo-referencing options based on Ground Control Points (GCP). PRISMA's fusion with other remote sensing data like LiDAR and SAR extends its applications, though its short lifespan and complex processing requirements limit its use. Future potential lies in classification applications using PRISMA hyperspectral images fused with LiDAR data and advanced machine learning models like deep CNNs [34]. FDD features include Q-based, high-frequency, and low-frequency types. The difference between standard Benford's law and extracted FDD features determines distribution differences. Pansharpening techniques such as SFIM, PCA, GSA, and GLP have been selected to improve image resolution. These techniques enhance LR image resolution using spatial filtering and are applicable to hyperspectral image fusion after band assignment [35]. The process for classifying urban tree species involves data preprocessing, tree crown extraction, shadow removal, and species classification using SVM, RF, and DenseNet. Four classification setups were tested: (1) WV2 VNIR bands alone, (2) WV2 VNIR with WV3 SWIR bands, (3) WV2 VNIR, WV3 SWIR, and LiDAR intensity, and (4) combining all with WV2 PAN bands. LiDAR-derived tree masks were applied to eliminate background and shadows, followed by a stratified thresholding technique. Tree crowns were isolated using a bimodal histogram threshold, and the non-vegetation background was removed based on spectral curve matching of the two examined bands [36].

3. Results and Discussion

The demonstrated approach for tree species classification and mapping uses spectral-spatial deep learning to map tree species proportions, accounting for the spatial resolution of modern satellite images and mixed forest complexities. It models predominant classes, species occurrences, and composition-basal area ratios and maximizes the use of forest inventory data for model creation. This approach is highly replicable and useful at large scales when inventory data and geo-referenced tree species proportions are available. The training dataset was based on the Forest Administration's forest parcel map and validated using independent data from regional forest inventory plots [2]. Poplar plantations are among the land cover types that S2 can map effectively in this study. The MLP algorithm produces annual statistics that traditional inventory methods cannot [12].

Table 5 summarizes various remote sensing studies, detailing their data sources, target areas, image resolution, classification methods, and achieved accuracies. It showcases diverse approaches, from traditional machine learning like Random Forest and logistic regression to advanced deep learning techniques such as U-Net++, CNNs, and 3D-CNNs. The studies cover different landscapes, including Wallonia (Belgium), the Padan plain (Italy), and boreal forests, utilizing various data sources like Sentinel-1 and 2, PlanetScope imagery, and LiDAR, with resolutions ranging from 0.5m to 20m. Overall, the table highlights the effectiveness of different classification methods in achieving varying levels of accuracy for diverse remote sensing applications.

Forest cover classification accuracy is comparable to results achieved using multi-seasonal Landsat Thematic Mapper-5 imagery and combined Sentinel-1 and Sentinel-2 datasets [15]. In a study utilizing airborne hyperspectral and LiDAR data, five widely used machine learning algorithms, RF, SVM GBM, ANN, and CNN, were evaluated for identifying four dominant tree species in a boreal forest. CNN models demonstrated the highest overall accuracy, with 3D-CNNs showing exceptional performance distinguishing coniferous species such as pine and spruce, highlighting their potential for forest industry applications [16]. Table 6 presents a comparative analysis of different classification algorithms applied to diverse remote sensing datasets. It lists the data sources (primarily Sentinel-1 and Sentinel-2), the specific classification methods used (including Support Vector Machine, Random Forests, k-Nearest Neighbour, and various Convolutional Neural Network architectures), and the resulting Overall Accuracies (OA). The table demonstrates how different algorithms perform on various datasets, highlighting the impact of data characteristics and methodological choices on classification accuracy, ranging from 30.6% to 98.1%, depending on the combination. Additionally, it includes experiments with different Sentinel bands and integrated datasets, along with comparisons to Very High Resolution (VHR) image analysis and temporal analysis.

Table 5. Parameter description of reference papers [2, 12-16]

Sources	Areas	Resolution	Methods	Accuracy
Sentinal-2 imagery	Wallonia region (Southern Belgium), the area covered 16,091sq km.	2.5m spatial resolution	Nested U-shaped neural network (UNet++) architecture	OA = 0.73 PA = 0.90 UA = 0.90
Sentinal-2 imagery	Padan plain in Northern Italy, the area covered 46,000sq km.	2.5m spatial resolution	Fully connected neural network (multilayer perceptron), Traditional logistic regression	MLP omission error rate = 2.77% ± 2.76% LR omission error rate = 8.9% ± 2.8%
Sentinal-2 (S2) and Planet Scope (PS)	Corine land cover of the two Belgian S2 tiles BeS and BeN together	10,20,2.5m spatial resolution	Convolution Neural Networks (CNNs) Residual-Learning Convolutional Neural Networks (RCNN)	OA = 98%
Sentinal-1 (S-1) and Optical Sentinal-2 (S-2)	National Forest Inventory (NFI) data, area covered 637,290 ha	-	Random Forest (RF) Algorithm LiDAR (light detection and ranging)	OA = 98% PA = 90% UA = 93%
Airborne data collection	Boreal forest ecosystem	0.5m spatial resolution	Deep learning (3D-CNN comparison with SVM, RF, GBM, ANN)	OA = 87%

Table 6. Parameter description of reference papers [3, 17, 19, 20]

Data Set	Classification Algorithm	Overall Accuracy
Sentinel-1 (S1) & Sentinel-2 (S2) in the Magdalena region within the Cundinamarca district in central Colombia	Integrated Sentinel radar VNIR layer 1. Support Vector Machine 2. Random Forests 3. k-Nearest Neighbour	1. OA=0.887 2. OA=0.555 3. OA=0.393
Sentinel-1 (S1) & Sentinel-2 (S2) in Serra de Monchique mountain in the southern region of Portugal, Algarve.	Random Forests 1.S1+S2 2. SI 3. S2	1. OA=0.981 2. OA=0.889 3. OA=0.974
Aerial photography in Selangor, Malaysia	1. Convolution Neural Network (CCN) 2. CNN + batch normalization 3. CNN + dropout 4. CNN + dropout + batch normalization	1. OA=0.932 2. OA=0.964 3. OA=0.958 4. OA=0.973
Sentinel-1 (S1) & Sentinel-2 (S2) in the Magdalena region within the Cundinamarca district in central Colombia	Sentinel-1A radar layer 1. Support Vector Machine 2. Random Forests 3. k-Nearest Neighbour	1. OA=0.306 2. OA=0.201 3. OA=0.169
Sentinel-1 (S1) & Sentinel-2 (S2) in the Magdalena region within the Cundinamarca district in central Colombia	Sentinel-2A VNIR layer 1. Support Vector Machine 2. Random Forests 3. k-Nearest Neighbour	1. OA=0.725 2. OA=0.481 3. OA=0.375
Sentinel-2 in Nice, Fran located on the eastern border with Italy.	RF and Geographic Object-Based Image Analysis (GEOBIA) 1. VHR (Very high resolution) image 2. Temporal analysis	1. OA=0.626 to 0.67 2. OA=0.743

Accurate LULC mapping faces challenges like class confusion, unclear boundaries, and low-resolution satellite images. Furthermore, multiple data sources, including temporal data, add complexity to the input layers. Identifying the most relevant features for classification is essential to effectively interpret satellite imagery layers [3]. GEOBIA, more than just a combination of segmentation and analysis techniques, is an evolving paradigm offering tools and guidelines for location-based research [18]. A key goal in monitoring ecosystem disturbance events is mapping vegetation cover, especially distinguishing between forest and non-forest vegetation before disturbances [19]. To prevent biased classification due to data imbalances, equal training and evaluation samples were used for all eight LCLU classes, though this may overlook class distribution ratios [21]. For example, forest and unclassified classes dominated, resulting in potential classification bias [1]. The GF2&S2-B10-9I model achieved 0.928 overall accuracy, demonstrating the effectiveness of the DeepLabV3 Plus model for complex land-cover classification [12].

In swamp vegetation, reduced spectral separability occurs due to fluctuating water levels [7], while deep learning models like a DCNN show better performance than traditional methods for water segmentation from satellite data [8]. Training strategies such as random and uniform sampling improved training efficiency and classification accuracy [10]. Segmentation models like FCN, U-Net, and DeepLabv3 achieved high performance with IoU and mIoU values of 83%-86% and 80%-84%, respectively [11]. NIR reflectance performed well for tree segmentation, with DLinkNet achieving 0.921 mJI [4]. Hybrid CNN effectively classified plant features using class activation mapping [23].

Sentinel-2 outperformed other sensors like RapidEye and Landsat 8 in vegetation prediction [24]. Hyperspectral data performed better than multispectral for Antarctic vegetation classification, especially in the SWIR range [25]. The RF algorithm showed the best accuracy for vegetation classification in coal mine reclamation areas [26]. Spectral models like SNIR and SRGB achieved high classification accuracy for tree species [22]. UNet, combined with transfer learning and residual networks, improved model performance for land cover classification [5]. Data augmentation and increased dataset size enhanced land cover mapping accuracy [28]. Fully convolutional networks with multi-resolution layers improved segmentation and learning efficiency [31]. SVM's performance depends on selecting optimal parameters like Gamma and C [6]. The fast iteration super-resolution algorithm outperformed SRCNN in terms of fusion performance and temporal complexity [32]. Hyperspectral data processing requires AI and machine learning expertise, with PRISMA data applicable across land and ocean applications [34]. DenseNet improved classification accuracy as more datasets were included, increasing from 75.9% to 82.6% overall [36].

4. Conclusion

The efficiency of spectral information is compared between UNet and DeepLabV3+ using evaluation metrics such as precision-recall curve, average precision, precision, recall, and F1 score. The mean Average Precisions (APs) for U-Net, SiU Net, and DeepLabV3+ were 0.921, 0.970 and 0.976, respectively, defining the performance of the reported methods. The models considered the correlation between RGB and NIR bands and used U-Net with distinct inputs. A reliable evaluation method was described for tree species proportion maps, assessing most species, composition, and proportions. Recent studies suggest incorporating time series data for tree species classification, leveraging phenology. The technique improves object detection in GEOBIA using VHR optical aerial images and multi-temporal Sentinel-2 data.

The Segment Anything Model (SAM) provided meaningful objects that can be classified more accurately with machine learning or CNNs, complementing SNIC clustering. Fusing deep-learning CNNs and GEOBIA algorithms can potentially improve LULC classification accuracy. DLinkNet outperformed other semantic segmentation models, particularly grasslands and tree occlusion scenarios.

Four models showed superior performance compared to state-of-the-art land use and cover classification methods, with future work targeting misclassification and accuracy improvement. High-resolution hyperspectral and multispectral imagery, especially from the WorldView-2 satellite, proved effective for monitoring mountain and coastal areas. Future research will explore applying deep learning to create benthic and vegetation maps and analyze changes using multi-temporal data. Also, land cover identification can be improved by determining the optimal number of encoders based on the spectral band's correlation coefficients.

References

[1] Won-Kyung Baek, Moung-Jin Lee, and Hyung-Sup Jung, "Land Cover Classification from RGB and NIR Satellite Images Using Modified U-Net Model," *IEEE Access*, vol. 12, pp. 69445-69455, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Corentin Bolyna et al., "Mapping Tree Species Proportions from Satellite Imagery using Spectral-Spatial Deep Learning," *Remote Sensing of Environment*, vol. 280, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Suzanna Cuypers, Andrea Nascetti, and Maarten Vergauwen, "Land Use and Land Cover Mapping with VHR and Multi-Temporal Sentinel-2 Imagery," *Remote Sensing*, vol. 15, no. 10, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Irem Ulku, Erdem Akagündüz, and Pedram Ghamisi, "Deep Semantic Segmentation of Trees Using Multispectral Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 7589-7604, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Furkat Safarov et al., "Improved Agricultural Field Segmentation in Satellite Imagery Using TL-ResUNet Architecture," *Sensors*, vol. 22, no. 24, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Javier Marcello et al., "Advanced Processing of Multiplatform Remote Sensing Imagery for the Monitoring of Coastal and Mountain Ecosystems," *IEEE Access*, vol. 9, pp. 6536-6549, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Man Liu et al., "Comparison of Multi-Source Satellite Images for Classifying Marsh Vegetation using DeepLabV3 Plus Deep Learning Algorithm," *Ecological Indicators*, vol. 125, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Kunhao Yuan et al., "Deep-Learning-Based Multispectral Satellite Image Segmentation for Water Body Detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7422-7434, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Md Osman Gani et al., "Multispectral Object Detection with Deep Learning," *International Conference on Computational Intelligence in Communications and Business Analytics*, Santiniketan, India, pp. 105-117, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Islombek Mirpulatov et al., "Pseudo-Labeling Approach for Land Cover Classification Through Remote Sensing Observations with Noisy Labels," *IEEE Access*, vol. 11, pp. 82570-82583, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Imran Ahmed et al., "Comparison of Deep-Learning-Based Segmentation Models: Using Top View Person Images," *IEEE Access*, vol. 8, pp. 136361-136373, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] G. D'Amico et al., "A Deep Learning Approach for Automatic Mapping of Poplar Plantations using Sentinel-2 Imagery," *GIScience & Remote Sensing*, vol. 58, no. 8, pp. 1352-1368, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Hamid Ghanbari et al., "A Meta-Analysis of Convolutional Neural Networks for Remote Sensing Applications," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3602-3613, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Nicolas Latte, and Philippe Lejeune, "Planet Scope Radiometric Normalization and Sentinel-2 Super-Resolution (2.5 m): A Straight Forward Spectral-Spatial Fusion of Multi-Satellite Multi-Sensor Images Using Residual Convolutional Neural Networks," *Remote Sensing*, vol. 12, no. 15, pp. 1-19, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Kristian Skau Bjerreskov, Thomas Nord-Larsen, and RasmusFensholt, "Classification of Nemoral Forests with Fusion of Multi-Temporal Sentinel-1 and 2 Data," *Remote Sensing*, vol. 13, no. 5, pp. 1-19, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Janne Mäyrä, et al "Tree Species Classification from Airborne Hyperspectral and LiDAR Data using 3D Convolutional Neural Networks," *Remote Sensing of Environment*, vol. 256, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Nicola Clerici, Cesar Augusto Valbuena Calderón, and Juan Manuel Posada, "Fusion of Sentinel-1A and Sentinel-2A Data for Land Cover Mapping: A Case Study in the Lower Magdalena Region, Colombia," *Journal of Maps*, vol. 13, no. 2, pp. 718-726, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[18] Thomas Blaschke et al., "Geographic Object-Based Image Analysis-Towards a New Paradigm," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 87, pp. 180-191, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[19] Giandomenico De Luca et al., "Integrated Use of Sentinel-1 and Sentinel-2 Data and Open-Source Machine Learning Algorithms for Land Cover Mapping in a Mediterranean Region," *European Journal of Remote Sensing*, vol. 55, no. 1, pp. 52-70, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] "Maher Ibrahim Sameen, Biswajeet Pradhan, and OmarSaud Aziz, "Classification of Very High-Resolution Aerial Photos Using Spectral-Spatial Convolutional Neural Networks," *Hindawi Journal of Sensors*, vol. 2018, no. 1, pp. 1-12, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[21] Abdulhakim Mohamed Abdi, "Land Cover and Land Use Classification Performance of Machine Learning Algorithms in a Boreal Landscape using Sentinel-2 Data," *GIScience & Remote Sensing*, vol. 57, no. 1, pp. 1-20, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[22] Andrew J. Chadwick et al., "Species Classification of Automatically Delineated Regenerating Conifer Crowns Using RGB and Near-Infrared UAV Imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[23] Yu-Kai Chen, Keng-Hao Liu, and Chinsu Lin, "Plant Species Recognition based on Hyperspectral Plant Canopy Images with Deep Learning," *IEEE International Geoscience and Remote Sensing Symposium*, Pasadena, CA, USA, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[24] Enoch Gyamfi-Ampadu, Michael Gebreslasie, and Alma Mendoza-Ponce, "Evaluating Multi-Sensors Spectral and Spatial Resolutions for Tree Species Diversity Prediction," *Remote Sensing*, vol. 13, no. 5, pp. 1-18, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] Junhwa Chi et al., "Spectral Characteristics of the Antarctic Vegetation: A Case Study of Barton Peninsula," *Remote Sensing*, vol. 13, no. 13, pp. 1-21, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[26] Xu Wang et al., "Analysis of Vegetation Canopy Spectral Features and Species Discrimination in Reclamation Mining Area Using in Situ Hyperspectral Data," *Remote Sensing*, vol. 16, no. 13, pp. 1-22, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[27] Qingyu Li et al., "Building Footprint Generation by Integrating Convolution Neural Network with Feature Pairwise Conditional Random Field (FPCRF)," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 7502-7519, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[28] Nadir Bengana, and Janne Heikkila, "Improving Land Cover Segmentation Across Satellites Using Domain Adaptation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol.14, pp. 1399-1410, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[29] Vladimir Iglovikov, Sergey Mushinskiy, and Vladimir Osin, "Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition," *arXiv Preprint*, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[30] Akmalbek Bobomirzaevich Abdusalomov et al., "Improved Feature Parameter Extraction from Speech Signals Using Machine Learning Algorithm," *Sensors*, vol. 22, no. 21, pp. 1-21, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[31] Long, Jonathan, Evan Shelhamer, and Trevor Darrell, "Fully Convolutional Networks for Semantic Segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440, 2015. [Google Scholar] [Publisher Link]

[32] Honglin Wu et al., "Remote Sensing Image Sharpening by Integrating Multispectral Image Super Resolution and Convolutional Sparse Representation Fusion," *IEEE Access*, vol. 7, pp. 46562-46574, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[33] Thomas Lacombea, Hugues Favrelierea, and Maurice Pilleta, "Modal Features for Image Texture Classification," *Pattern Recognition Letters*, vol. 135, pp. 249-255, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[34] Riyaaz Uddien Shaik, Shoba Periasamy, and Weiping Zeng, "Potential Assessment of PRISMA Hyperspectral Imagery for Remote Sensing Applications," *Remote Sensing*, vol. 15, no. 5, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[35] Xiankun Hao et al., "A No-Reference Quality Assessment Method for Hyperspectral Sharpened Images via Benford's Law," *Remote Sensing*, vol. 16, no. 7, pp. 1-32, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[36] Sean Hartling et al., "Urban Tree Species Classification Using a WorldView-2/3 and LiDAR Data Fusion Approach and Deep Learning," *Sensors*, vol. 19, no. 6, pp. 1-23, 2019. [CsrossRef] [Google Scholar] [Publisher Link]