International Journal of Engineering and Geosciences
https://dergipark.org.tr/en/pub/ijeg
e-ISSN 2548-0960

# Evaluating the performance of deep learning-based segmentation algorithms for land use land cover mapping in a heterogenous vegetative environment

**İskender Berkay Sür[1], Ugur Alganci [*1], Elif Sertel [1]**

[1] *Istanbul Technical University, Geomatics Engineering Department, Istanbul, Türkiye, sur20@itu.edu.tr, alganci@itu.edu.tr, sertele@itu.edu.tr*

Cite this study: Sür, İ.B., Alganci, U., & Sertel, E. (2025). Evaluating the performance of deep learning-based segmentation algorithms for land use land cover mapping in a heterogenous vegetative environment. International Journal of Engineering and Geosciences, 10(3), 380-397.

https://doi.org/26833/ijeg.1528938

**Keywords**
Land Use - Land Cover
Unet++
DeepLabv3+
Worldview 3

**Research Article**

Received:06.08.2024
Revised: 14.12.2024
Accepted:10.03.2025
Published: 01.10.2025

**Abstract**

Land Use and Land Cover (LULC) maps are important geospatial information sources for different applications such as city planning, vegetation analysis, natural resource management, natural disaster analysis, and land change determination. In recent decades, the demand for more frequent creation and updating of LULC maps has grown significantly, driven by the rapid and continuous changes occurring on the Earth surface. Moreover, the increased availability of satellite images and processing power led to improvements in LULC mapping. However, traditional classification approaches are prone to several errors emerging from high human interaction and algorithm limitations. In addition, they generally suffer from processing time performance due to software limitations and generally singular hardware configurations, especially when very high resolution (VHR) images are of concern. In this study, we aim to produce LULC maps of the Aksu region of Bursa city Türkiye, using Worldview-3 VHR images and deep learning (DL) methods. We applied two widely used DL architectures, Unet++ and DeepLabv3+, and evaluated results using overall accuracy, average accuracy, error matrix, weighted accuracy, recall, precision, F-1 score, IoU score, and kappa metrics. Among several experimental setups, we achieved the best accuracy with the Unet++ architecture, using the ResNeXt-50 backbone and Adam optimizer, resulting in an approximately 84% IoU score and 91% F-1 score. This study demonstrates that utilizing appropriate datasets and CNN-based segmentation models for LULC mapping ensures efficient, accurate, and high-performance results, significantly contributing to long-term monitoring and sustainable development goals.

## 1. Introduction

The production and revision of Land Use and Land Cover (LULC) maps are critical in many fields currently. Accurate and timely LULC maps are required for a variety of applications, including urban and regional planning, natural resource and environmental management, disaster and hazard monitoring, and food security [1-3]. The use of these maps in such diverse and important areas also requires them to be up-to-date [4-7].

According to Carter and Herold (2019), while national land cover maps produced by certain countries may be of high quality in specific contexts, their accuracy and update frequency may be suboptimal [8]. This demonstrates the need for more effective and faster approaches to update LULC maps in a world that is continually and rapidly changing as a result of human and technological influences. Update requirement of LULC maps can occur frequently in some landscapes, such as industrial zones, as a result of variables such as unstable agricultural markets and an increase in population owing to migration [9].

LULC maps are also very important in sustainability efforts. The "Brundtland Report" introduced the concept of "Sustainable Development" in 1987, which necessitated a rethinking of the main principles of future planning. It is well known that LULC maps are crucial for monitoring and continuing indicators for the Sustainable Development Goals (SDG). National statistics and geospatial information can be efficiently used to track SDG indicators by combining appropriate satellite and Earth observation datasets, such as land cover maps [8].

The primary goal of LULC map classification is to identify the physical land use and land cover types within a region [10]. The data required for producing LULC maps can be obtained from various sources. Among them, the most popular and widely used one is the

remote sensing (RS) based images. According to Shi et al. (2019), various organizations, such as the European Space Agency (ESA), the National Aeronautics and Space Administration (NASA), the United States Geological Survey (USGS), and the National Oceanic and Atmospheric Administration (NOAA), provide extensive free RS and geospatial data for LULC mapping. In addition, social media platforms such as Facebook, Twitter, and Instagram can now generate large amounts of data with geospatial information that can be used for LULC mapping [11]. Even though RS images and the above-mentioned data are used for LULC maps, these data must be of certain standards, accuracy, quality, and resolution, to generate accurate maps for its intended purpose. In the field of RS, there is access to a large amount of data, but most of the data is unlabeled, therefore, a classifier is needed to segment this data [12]. Although several datasets are freely available for different deep-learning tasks, it is still problematic to have reliable references or ground truth data, especially for input images [13].

Recently, machine learning (ML)based algorithms were tested and used for LULC classification. One of the recent studies evaluated different ML based algorithms in LULC mapping on heterogenous environment that is defined by 11 classes. Authors of the study evaluated logistic regression, k-nearest neighbor, random forest and support vector machine classifiers with different image setups, and concluded with higher performances of random forest and SVM classifiers applied on bi-temporal satellite images [9]. Currently, deep learning (DL) models use architectures based on multi-layer convolutional neural network (CNN) architectures. Henry et al. (2019), define a neural network as a mathematical abstraction loosely based on the functioning of nerve cells (neurons) in the brain. A neural network is made up of layers of neurons, and the term deep learning refers to a neural network that has many more layers and parameters than a standard neural network. These algorithms learn to perform tasks by identifying patterns in large, labeled (classified) data sets [14]. DL models have become a common approach for analyzing satellite imagery. More recently, developments in DL-based architectures tailored for the field of RS research have led to quicker and more precise LULC mapping, particularly for segmentation and classification needs [13].

The current CNN architectures, AlexNet, VGGNet and GoogLeNet, have achieved significant success in image segmentation tasks and based on these great successes, Long et al. proposed to adapt modern DESA architectures to Fully Convolutional Neural Networks (CNNs) for use in semantic segmentation [15]. DESAs have been successfully adapted to solve remote sensing problems [14]. Considering all these achievements, a few performance-based CNN architectures that stand out are VGGNET, ViT, DenseNET, ResNET, UNet, DeepLabv3+ and GoogLeNET. According to Sertel et al., the DeepLabv3+ architecture gave the highest accuracy results in the generation of LULC maps among Pyramid Attention Network (PAN), U-Net++, Feature Pyramid Networks (FPNs), Linknet, Pyramid Scene Parsing Network (PSPNet) architectures with ResNeXt50 backbone [13]. DeepLabv3+ with ResNeXt50 encoder was chosen in light of its superior performance in semantic segmentation tasks (especially when processing complex boundaries and multiple classes) according to Sertel et al. study. This model is an excellent option for mapping land use and land cover in VHR satellite imagery due to its proven flexibility in various geographic contexts [13]. In their thorough analysis of the image scene classification task, Cheng et al. [56] covered the specifics of image scene classification techniques based on autoencoders, convolutional neural networks (CNNs), and generative adversarial networks (GANs). A thorough review of the semantic segmentation of remotely sensed images was carried out by Yuan et al. [57]. They covered the CNN architectures used in semantic segmentation, including DeepLab, SegNet, and U-Net, in particular.

ResNet50 and its extended version, ResNeXt50, are utilized as backbones to examine their impact on feature extraction in segmentation tasks. The investigation focuses on evaluating ResNet's capability to mitigate overfitting in deep architectures and ResNeXt's potential to enhance performance through its wider network structure enabled by group connections. Sertel et al. obtained 0.8946 IoU and 0.9434 F-1 scores for ResNeXt50 and 0.8732 IoU and 0.9308 F-1 scores for Resnet50 with DeebLabv3+ in their segmentation with VHR images and completed segmentation with high accuracy [13].

Segmentation involves breaking an image into smaller parts and characterizing each uniformly, which is crucial for constructing LULC maps using deep learning (DL) structures. This process, combined with classification, helps define homogeneous land cover zones. Semantic segmentation faces a conflict between local (where) and global (what) information [15]. Architectures like DeepLabv3+, UNet, Mask R-CNN, SegNet, and Fully Convolutional Networks (FCN) employ encoding and decoding structures for this purpose [16, 17]. Convolution processes create new output values by combining data points with their neighbors and are used for operations such as edge detection, blurring, and sharpening. The stride determines the shift of the convolution process, and the filter size affects the area of application. Larger filter sizes and fewer steps increase processing time and load.

High segmentation accuracy of RS data often requires a large collection of training and validation samples [18]. VHR data sets with a very high spatial resolution, enough images, and different class types are needed for the production of LULC maps. The usage of segmentation and image analysis techniques has expanded and become easier owing to VHR data. Different classes can be detected in such data effectively, and object-based segmentation techniques outperform pixel methods at high spatial resolution. One of the primary sources for creating VHR datasets is RS satellites, which are currently in widespread usage and were constructed using cutting-edge technologies. VHR images can be received from the WorldView 3, GeoEye 1, Pleiades, and Skysat satellites, which are some examples

that are currently in service. Additionally, these satellites provide a range of spectral resolutions.

Pixel-based and object-based segmentation are two examples of segmentation types. Considering the integrity and properties of the objects in the image, each pixel is allocated to a particular object class. According to recent studies, object-based image analysis (OBIA) is an essential component of image segmentation, particularly when it comes to identifying objects and blending geographic characteristics [19]. Apart from object-based segmentations, there are also pixel-level segmentations. This segmentation method provides segmentation according to the reflectance value of each pixel.

This study aims to produce accurate LULC map of the Aksu region of Bursa city Türkiye, using VHR satellite imagery and DL approaches. The performance of two CNN-based segmentation architectures, effects of loss functions, backbones, batch size, learning rate, and iteration size were evaluated in different experiments. Different accuracy metrics are presented to compare the performance of various hyper-parameters for the Unet++ and DeepLabv3+ segmentation architectures.

## 2. Method

### 2.1. Study area and dataset

The Aksu region of Bursa province in Türkiye was chosen as the case study area, since it contains a variety of agricultural and forest areas, as well as land use areas such as urban settlements and mining (Figure 1). This region is located at 40.18°N, 29.07°E, within the boundaries of Bursa, at an elevation of 150 m.

The images used in this study were acquired by Worldview 3 (WV-3) satellite and collected on September 6, 2020. The four multi-spectral bands (red, green, blue, and near infrared) at 1.24 m resolution and the panchromatic (PAN) image at 31 cm resolution were combined using the PANSHARP2 algorithm to create pan-sharpened images at 30 cm spatial resolution with four spectral bands [13, 20, 21].

In a previous study, the same pan-sharpened (PSP) WV-3 images of the Aksu and Kestel regions were segmented and classified using an object-based approach (OBIA) using the E-cognition software [13]. We used those classified images as labelled date. This dataset includes eight second level CORINE classes, including forest, mining-casting and construction sites, road and railway, discontinuous urban texture, arable farmland, heterogeneous agricultural areas, permanent agricultural areas, inland water resources, and areas with no data.

Since the distribution of each class in the images is unbalanced (Figure 2), the unrepresented classes were adjusted in each sample to ensure proper training during the DL model implementation. For this purpose, the "compute_sample_weight" function from Ski-learn was used to calculate the weights of each sample, taking into account the number of different classes in each sample [22,23].

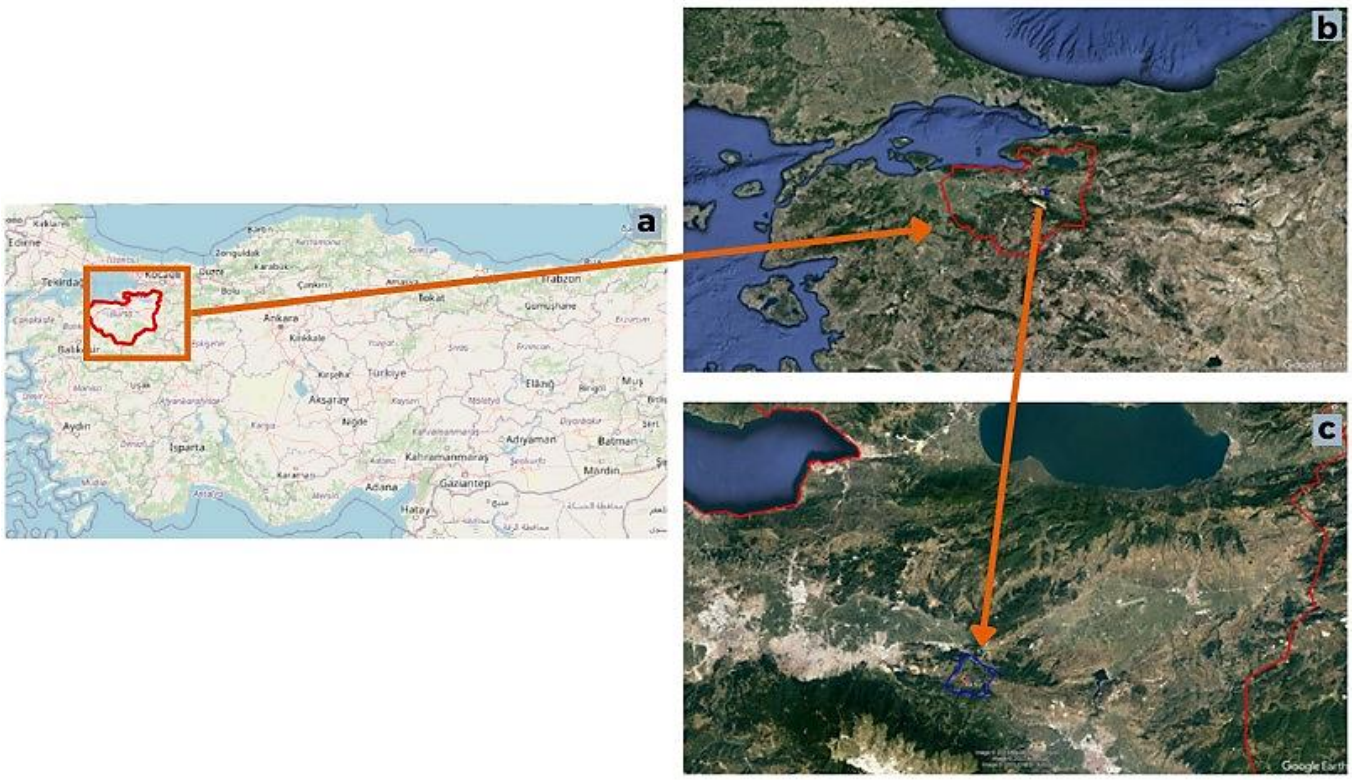


**Figure 1.** a) General representation of the Bursa region within the borders of Turkey through the Open Street map, b) Satellite image representation of the Bursa province from Google Earth, c) Satellite image representation of the Aksu region from Google Earth.

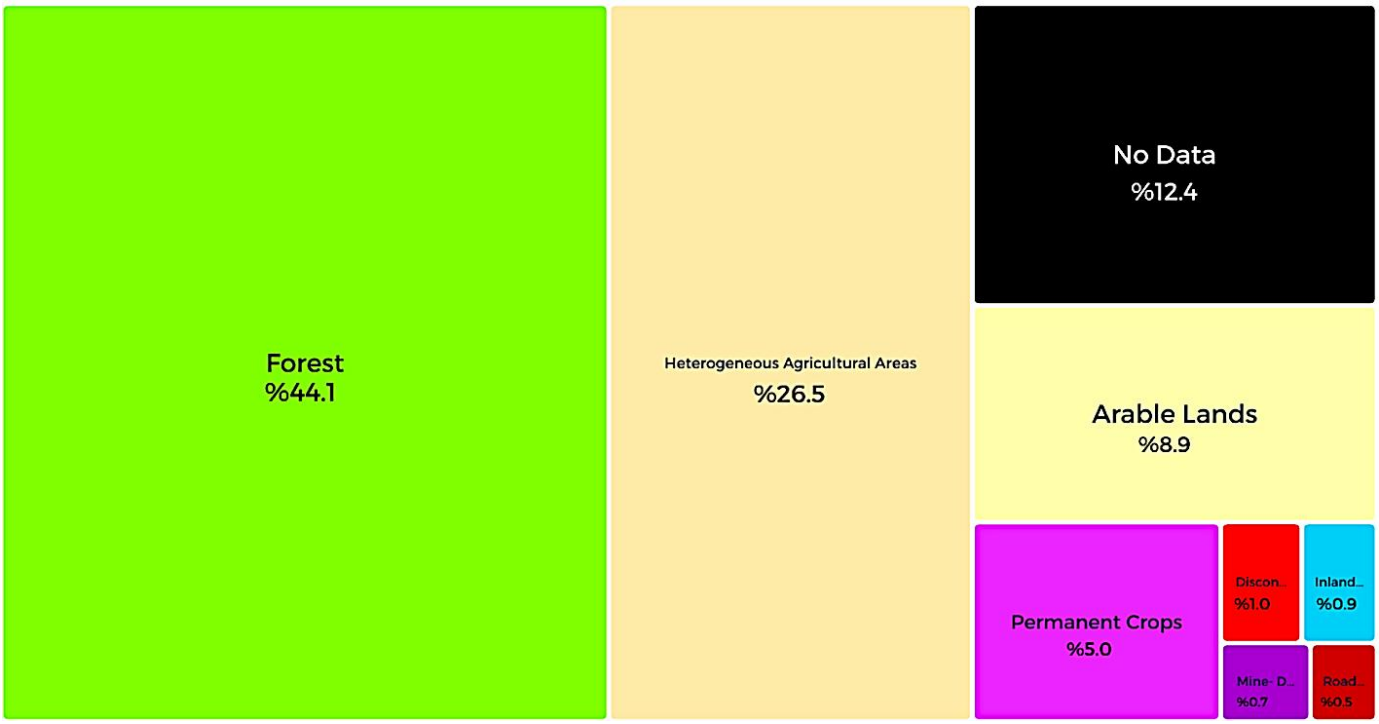


**Figure 2**. Percentage representation of LULC class distributions in the Aksu region with the tree map graphic method [13].

Figure 2 shows the distribution of LULC classes: Forests are 44.1%, heterogeneous agricultural areas are 26.5%, no data is 12.4%, arable agricultural lands are 8.9%, permanent agricultural areas are 5.0%, intermittent urban texture is 1.0%, internal water resources are 0.9%, mine-casting, and construction sites are 0.7%, roads, and railways are 0.5%. There is a total of 586 image patches of 512*512 pixel size in the Aksu WV-3 dataset. When developing statistical and machine learning models, it is common to divide the data set into two parts, training and testing [24,25]. On the other hand, the validation set can be used to fine-tune the model performance, such as selecting hyperparameters or regularization parameters in the model [26]. Based on this knowledge, the dataset was divided into 70% training, 20% validation, and 10% test, thus, 412 patches for training, 116 patches for validation, and 58 patches for testing. The sample view of the dataset with true color images and ground truth pairs is provided in Figure 3.

To increase the volume of the data set, augmentation techniques are adopted by applying basic image processing techniques [12,13]. The albumentations library was used to apply these techniques, including padding, random rotation, mirroring, shifting, scaling, rotation, and normalization, to increase the robustness of the model against various image deformations and orientations, minimize overfitting, and improve performance.

## 2.2. Methodological Structure

After the generation of the data set, we used the Pytorch library to utilize DL models. PyTorch is an open-source machine learning library used for applications such as computer vision and natural language processing. It was mainly developed by Facebook's artificial intelligence research laboratory. As a Python-based library, PyTorch can perform instantaneous dynamic tensor operations with automatic differentiation and GPU support, while exhibiting a performance comparable to the fastest DL libraries [27].

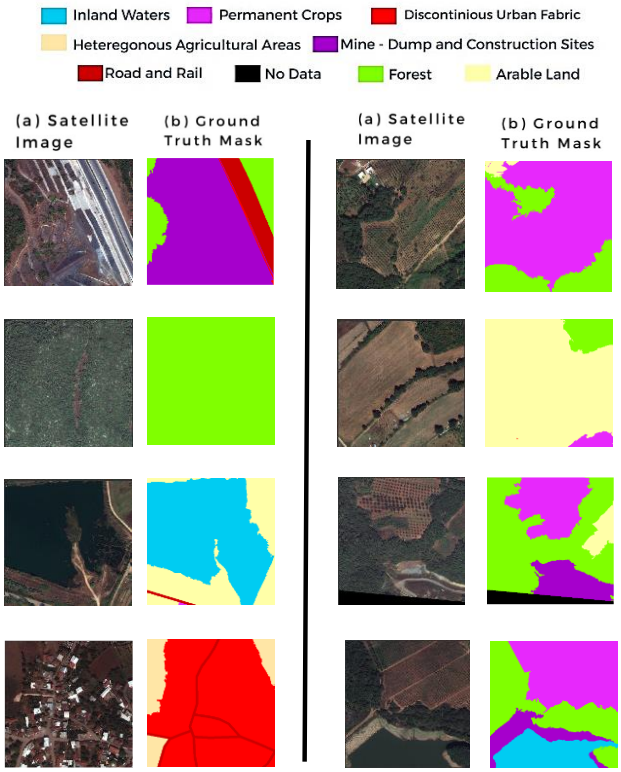


**Figure 3**. RGB true color satellite images and ground truth masks of the Aksu region from the training dataset.

Several experiments were conducted based on optimization methods, loss functions, decoders, and backbones. We implemented Unet++ and DeepLabv3+ segmentation architectures, with ResNeXt50 and

Resnet50 backbones and Adam and RMSprop optimizers. Furthermore, the impacts of Dice-Focal and Focal loss configurations were also tested. The Adam optimizer combines the advantages of momentum and RMSprop. While RMSprop is memory-efficient and can yield more effective results with larger batches on platforms like Kaggle by continuously updating the learning rate at each step, a momentum-based optimizer typically requires more memory but provides more stable and faster learning [28].

The learning rate during training is a hyperparameter that determines how much the weights of the model will be changed in each training network. In this study, a learning rate of $10^{-5}$ was selected to prevent overfitting and ensure more regular and stable model training. Additionally, the softmax activation function, commonly used in multi-class segmentation tasks, was employed to obtain class probabilities for each pixel, ranging from 0 to 1. This ensures that the sum of all probabilities equals 1, allowing the image to be segmented into regions corresponding to different classes [29,30].

The batch size defines the amount of data that can be processed within the GPU memory (VRAM). Larger batch sizes can accelerate training by allowing more data to be processed simultaneously. However, as the batch size increases, GPU memory usage also increases, which may lead to memory limitations. Therefore, the GPU's memory capacity must be considered when selecting an appropriate batch size. Additionally, excessively large batch sizes can negatively impact the stability of the model. This study was conducted on the Kaggle platform using a P100 GPU [31]. As a result, the batch size was set to 4 for each training session, as larger batch sizes would trigger memory warnings and potentially reduce training accuracy.

During each epoch, the model processes all the training data once, learning from it and improving incrementally. As the number of epochs increases, the model sees the data more frequently, typically leading to better performance. For the P100 GPU with 16 GB of memory used in Kaggle platform, the optimal number of epochs for achieving high-accuracy results was determined to be 100, based on both the platform's time constraints and experimental findings.

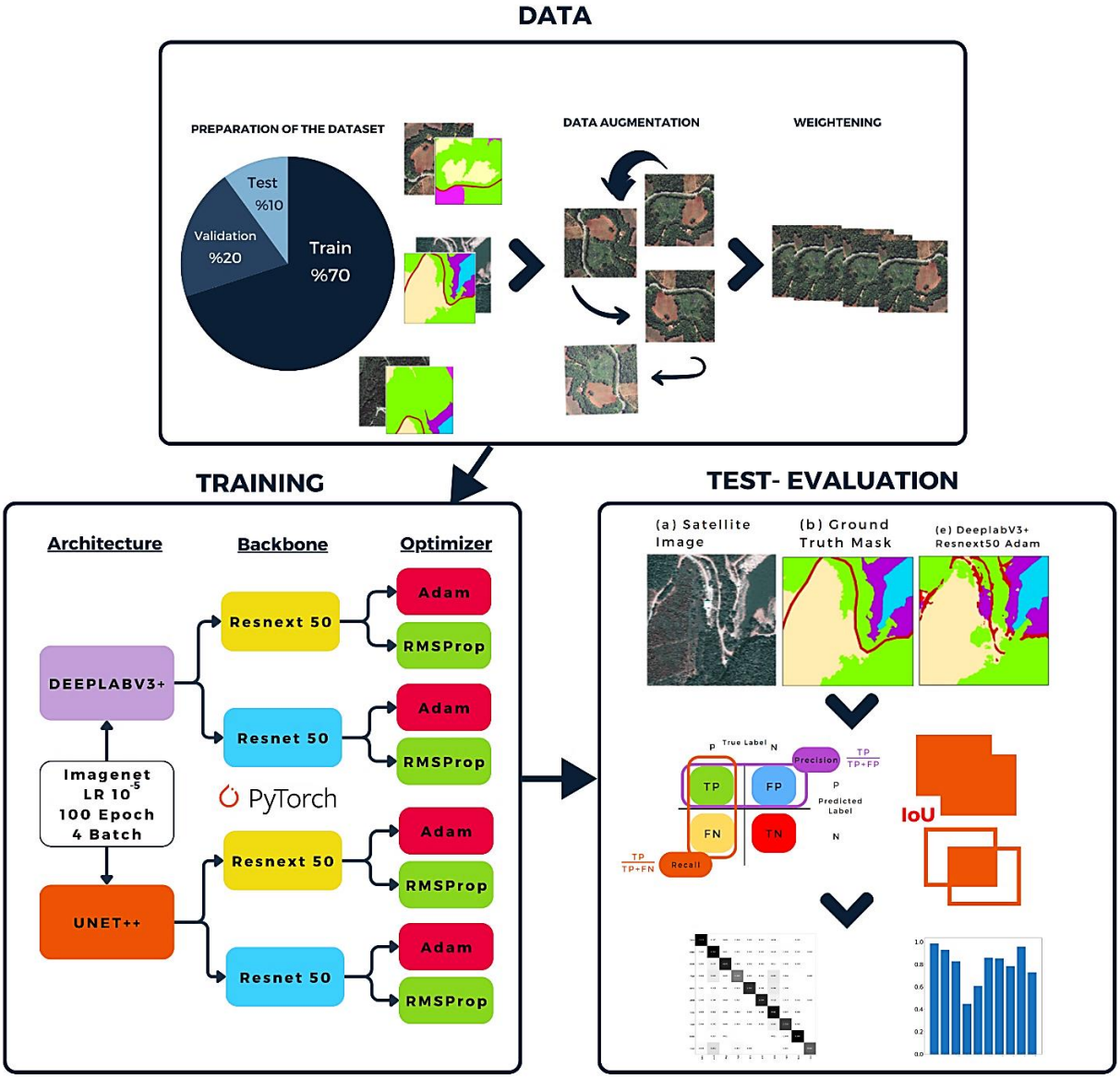


Figure 4. General flowchart of the study.

### 2.3. Segmentation Architectures

Various machine learning techniques have been evaluated in previous studies, with results indicating that Convolutional Neural Networks (CNNs) outperform non-deep learning methods across multiple datasets [32-35]. To recognize features, CNNs use a segmentation approach to interpret them. A typical CNN architecture is shown in Figure 4. Semantic segmentation of digital images involves assigning a class label to each pixel in the image (e.g., road, grass, forest). The goal is to group pixels that belong to the same perceptual category, thereby providing contextual meaning to the pixels [36]. Standard CNNs consist of convolutional layers and pooling layers, followed by fully connected layers with classifiers, utilizing activation functions to introduce non-linearity [37].

The DeepLab is a deep CNN model that is used to perform semantic labeling on a pixel-by-pixel basis. DeepLab addresses challenges such as reduced feature resolution, the presence of objects at multiple scales, and limited spatial accuracy [38]. The latest version of the DeepLab architecture used in this study, which is DeepLabv3+, has an advanced decoder architecture that allows for extensive refining of object boundaries via Atrous convolutions and the incorporation of a decoder module, all while keeping DeepLabv3+'s multi-scale contextual capabilities. The proposed DeepLabv3+ model builds on DeepLabv3 by using Atrous convolution, which can modify the density of encoded features as much as computational resources allow, and an efficient solver module that refines detailed object boundaries [39].

The second architecture evaluated in this study is Unet, which uses an encoder and decoder components. The encoder learns the image's features while lowering its size, whereas the decoder uses the learned features to produce a higher resolution output via up-sampling and linking, as well as the option to switch between links. UNet++ consists of several U-Net architectures of different depths, and their decoders are densely connected at the same resolution with redesigned jump links, enabling the gradual fusion of image features across the network horizontally and vertically [40].

### 2.4. Deep Learning Backbones in Segmentation Models

Backbone is the name given to the layers that form the basic building block of the model in DL, especially in image processing applications. These layers are used to extract meaningful features from the input image. The backbone, usually consisting of convolutional neural networks, forms the first part of the model and produces rich feature maps for the following layers. Thus, the backbone serves as the core feature extraction mechanism of the model. As deeper networks start to converge, a degradation problem arises because accuracy levels reach saturation and then drop rapidly as network depth increases [41]. Overfitting is not associated with such degradation, and training error increases as more layers are added to an appropriately deepened model [42,43].

A deep residual learning framework was developed to ensure that each of the stacked layers conforms to a residual mapping precisely. This approach avoids assuming that all these layers will immediately conform to the intended base mapping [41]. ResNeXt belongs to the ResNet (Residual Network) family, a type of deep neural network. Besides the ResNet architecture, ResNeXt also includes a dimension called cardinality. In addition to depth and breadth, the size of the transformation set, called cardinality, is considered a fundamental factor in the model [35]. ResNeXt50 indicates a 50-layer ResNeXt model. ResNeXt50 is often used in tasks such as image segmentation, detection, and segmentation and serves as the backbone of the model.

### 2.5. Activation and Loss functions and optimizers

Many factors influence the performance of DL architectures and analysis accuracy. Epochs are the periods during which the neural network model processes all training data. The model can be trained gradually, a process known as training phases. Some functions help establish the model's output limitations and accuracy while also accelerating learning. These are known as activation functions. Common functions include Softmax, Relu, tanh, elu, sigmoid, and leaky Relu. The Softmax, Relu, and Elu functions are among the most popular and useful in providing accurate results. The Softmax function computes the likelihood of each class across all classes for a given pixel [37].

The loss function evaluates the performance of the model by measuring the difference between its predictions and the actual values. In model training, the loss is minimized. Dice loss tries to maximize the overlap between the predicted results of the segmentation model and the actual results. That is, Dice loss is used as a complement to the Dice score coefficient to improve the model performance and is usually expressed by Equation 1 as the Dice score coefficient. The Dice score coefficient is a commonly used overlap measure to evaluate segmentation performance when gold standard or real data is available [44]. Focal loss allows the model to focus on difficult samples by emphasizing the minority class when there is a class imbalance. Focal loss can handle extreme foreground-background class imbalances [45]. Dice-Focal loss combines the advantages of both losses. In the Dice-Focal loss equation (1) the first term represents the Dice loss and the second term is the Focal loss weighted by a coefficient of 0.5, with the operator γ acting as a relaxation parameter that adjusts how important correctly or misclassified samples are. In the Dice loss function, $p_i$ and $g_i$ represent the matched pixel values of the predicted and actual values, respectively, while the $a_t$ term in the Focal loss function is a weight-hyperparameter offset that scales the main term to solve the class imbalance problem [13,46,47]. Optimizers refer to algorithms that manage the model's learning process by determining how to update the model parameters. Models usually try to learn from data by minimizing a loss function. The optimizer determines

how to adjust weights and biases to reduce this loss. Optimizers thus have a critical role in ensuring that the model learns efficiently.

$$L = \frac{2\Sigma_i^N p_i g_i}{\Sigma_i^N p_i^2 + \Sigma_i^N g_i^2} + ( -a_t(1-p_t)^\gamma loglog(p_t)) \times 0.5 \quad (1)$$

Adagrad optimizer uses customized learning rates for each parameter. This is useful for sparse data. The RMSprop optimizer prevents Adagrad from rapid weight reduction. It only takes into account the last few updates. Adagrad is an algorithm that customizes gradient-based optimization by applying different learning rates for rare and frequent parameters, making it suitable for sparse data [48]. Adadelta is an improved version of Adagrad and compensates for its weakness by keeping past gradients in a limited window to keep the learning rate constant [49]. Adaptive Moment Estimation, or Adam optimizer for short, provides adaptive learning rates by combining momentum and scaled gradients. It uses first and second-moment estimates. The Adam optimization algorithm provides faster and more effective optimization by using both the gradient direction with the momentum method and the decreasing averages of past gradients [50]. Different optimizers can be used for various purposes and considering factors such as memory, speed, and parameters.

### 2.6. Evaluation metrics

LULC maps are evaluated using different metrics such as, overall accuracy, average accuracy, error matrix, weighted accuracy, recall, precision, F-1 score, and kappa statistics [51]. F-1 score represents the harmonic mean of the precision and recall scores, which measure the accuracy and recall of the classifier. Having unbalanced precision and recall scores will result in a low F-1 score while having balanced precision and recall scores will ensure a higher F-1 score [13]. The numbers of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) are used to calculate the values of these metrics [52]. Figure 5 shows the calculations for recall and precision.

Congalton and Green state that, analyzing the reasons for the differences in the confusion matrix may be one of the most important steps when creating a LULC map with RS data [53]. Confusion matrices show the rate of agreement of the classes obtained as a result of the experiment with reality and the rate of confusion with other classes [55]. IoU is a common metric that compares the similarity between two shapes in a scale-independent manner by converting their shape properties such as width, height, and position into a normalized measure [54].

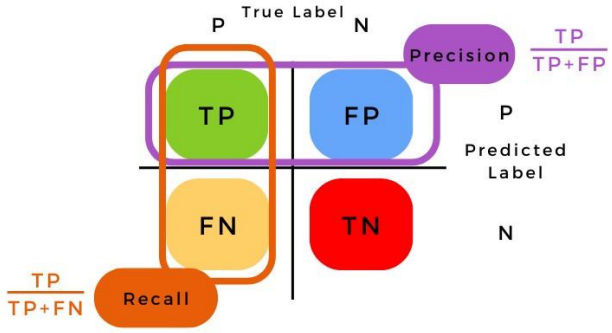


**Figure 5**. Illustration of True positive, False positive, False negative and True negative metrics calculation.

## 3. Results

In DL-based LULC mapping of the Aksu region, the main aim was to keep the value of the loss function as low as possible during the training of the model to provide predictions that are as close to real-world data as possible. In addition, evaluation with metrics were employed to analyze and enhance the model's performance. Dice was assessed independently in terms of Focal loss using both metrics and visual interpretation. Among several setups, the eight best-performing experiments were presented in this paper that are:

(c) DeepLabv3+ Resnet50 Adam,
(d) DeepLabv3+ Resnet50 RMSprop,
(e) DeepLabv3+ ResNeXt50_32x4d Adam,
(f) DeepLabv3+ ResNeXt50_32x4d RMSprop,
(g) Unet++ Resnet50 Adam,
(h) Unet++ Resnet50 RMSprop,
(i) Unet++ ResNeXt50_32x4d Adam,
(k) Unet++ ResNeXt50_32x4d RMSprop

### 3.1. Statistical Evaluation

Table 1 displays the F-1 score values for each class, while Figure 6 depicts a graphical representation of these data for experiments with Dice-Focal loss. As a result, the values with the greatest F-1 score for Dice-Focal correspond to forest and inland water resources, as well as agricultural areas. At the same time, the lowest and most imperceptible class is road and rail. As shown in Table 2 and Figure 7, the accuracy values of the models indicate that experiments using the Deeblabv3+ architecture produced better results in terms of IoU, F-1 score, precision, and accuracy metrics. According to the results, although the experiments using the Adam optimizer give better results compared to RMSprop, it is seen that RMSprop gives high-accuracy results with the Deeplabv3+ architecture. Hence, the best experiment is the one with RMSprop.

It can be seen that Unet++ outperforms Deeplabv3+ in terms of Dice-Focal loss. However, the best experimental result was achieved with Deeplabv3+ architecture, ResNeXt 50 backbone, and RMSProp when evaluated with Dice-Focal loss based on 0.84 IoU and 0.91 F-1 Score. When assessing the results on the test dataset, it is evaluated that some classes are close to the real ground mask with high accuracy. In contrast, some classes with particularly similar spectral reflectance are mixed. In addition, classes with a lower amount of

training data provided low accuracy despite weighting with the scikit-learn function in some experiments. The heterogeneous agricultural areas class was the most difficult class to distinguish among the classes due to mixing with the discontinuous urban texture and the road and railway class.

When the experiments with Focal loss are evaluated in terms of accuracy metrics (Table 3 and Figure 8), the experiments with the Unet++ architecture yielded higher performance in terms of IoU, F-1 Score, precision, and accuracy. According to comparative evaluation, it is seen that the ResNeXt 50 backbone with Unet++ architecture gives higher accuracy results independent of the optimizer. In terms of Focal loss, all experiments are close to each other. The best-performing experiment was the setup consisting of Unet++ architecture, ResNeXt 50 backbone, and Adam optimizer with 0.85 IoU and 0.92 F-1 Score for Focal loss.

In addition, similar to the Dice-Focal loss experiments, the heterogeneous agricultural areas class was the most difficult class to distinguish among the classes by mixing with the discontinuous urban texture and the road and railroad classes. Accordingly, the values with the highest F-1 Score for Focal loss are again for forest and inland water resources and agriculture-related areas. At the same time, the lowest and imperceptible classes are road and railroad.

**Table 1.** Class-wise F-1 Score values of different experiments with Focal Loss

| Experiment / Class | No Data | Forest | Mine-Dump Construction Sites | Road and Rail | Discontinuous Urban Fabric | Arable Land | Heterogonous Agricultural Areas | Permanent Crops | Inland Waters |
|---|---|---|---|---|---|---|---|---|---|
| (c)DeepLabv3+ Resnet50 Adam | 0.9820 | 0.9318 | 0.8115 | 0.4752 | 0.7172 | 0.8602 | 0.8689 | 0.7746 | 0.9654 |
| (d)DeepLabv3+ Resnet50 RMSprop | 0.9825 | 0.9330 | 0.8516 | 0.3451 | 0.7543 | 0.8751 | 0.8592 | 0.7789 | 0.9746 |
| (e)DeepLabv3+ ResNeXt50_32x4d Adam | 0.9823 | 0.9312 | 0.7725 | 0.4239 | 0.8248 | 0.8582 | 0.8569 | 0.7807 | 0.9720 |
| (f)DeepLabv3+ ResNeXt50_32x4d RMSprop | 0.9827 | 0.9395 | 0.8291 | 0.4896 | 0.7749 | 0.8499 | 0.8852 | 0.8253 | 0.9685 |
| (g)unet++ Resnet50 Adam | 0.9865 | 0.9310 | 0.7883 | 0.5941 | 0.8281 | 0.8227 | 0.8520 | 0.7753 | 0.9631 |
| (h )Unet++ Resnet50 RMSprop | 0.9821 | 0.9280 | 0.8165 | 0.5814 | 0.7107 | 0.8061 | 0.8374 | 0.7750 | 0.9541 |
| (i)Unet++ ResNeXt50_32x4d Adam | 0.9853 | 0.9358 | 0.7057 | 0.5727 | 0.7302 | 0.8522 | 0.8836 | 0.8125 | 0.9629 |
| (k)Unet++ ResNeXt50_32x4d RMSprop | 0.9858 | 0.9203 | 0.8229 | 0.4897 | 0.7573 | 0.6824 | 0.7864 | 0.5820 | 0.9728 |

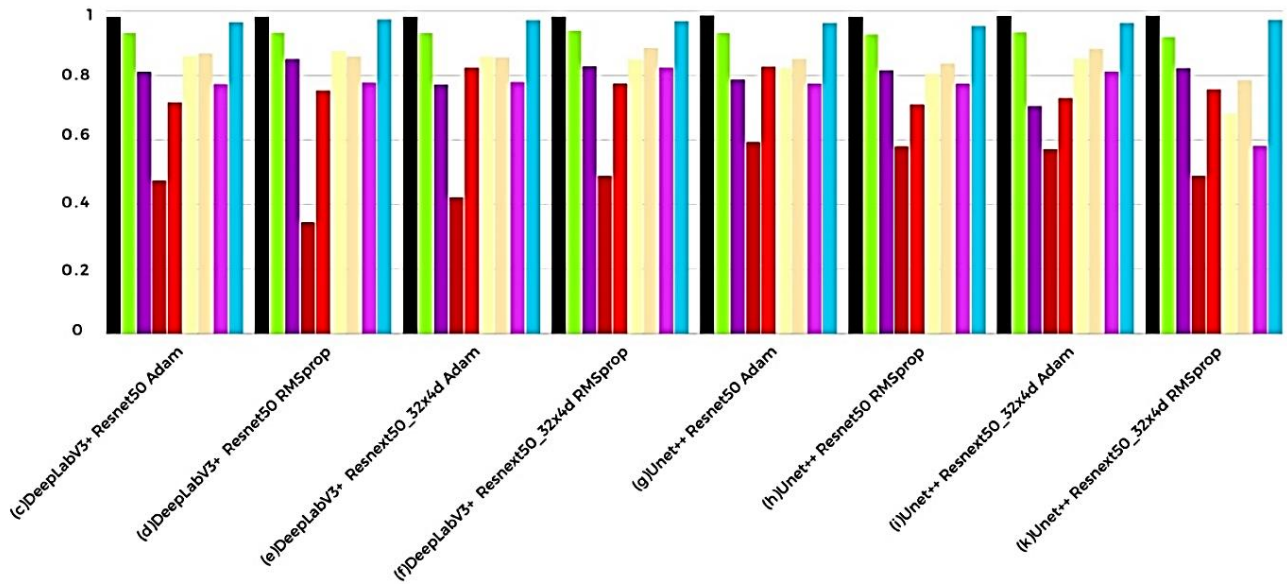


**Figure 6.** Graph visualization of the final class-based F-1 scores for each experiment.

**Table 2.** Metric Evaluations of Tested Models of Dice-Focal Loss Experiments.

| Experiment / Evaluation Metric | Dice-Focal Loss | IOU Score | F-1 Score | Precision | Recall |
|---|---|---|---|---|---|
| (c)DeepLabv3+ Resnet50 Adam | 0.5682 | 0.8290 | 0.9037 | 0.9040 | 0.9034 |
| (d)DeepLabv3+ Resnet50 RMSprop | 0.5682 | 0.8289 | 0.9036 | 0.9042 | 0.9029 |
| (e)DeepLabv3+ ResNeXt50_32x4d Adam | 0.5685 | 0.8233 | 0.9000 | 0.9001 | 0.9000 |
| (f)DeepLabv3+ ResNeXt50_32x4d RMSprop | 0.5673 | 0.8411 | 0.9112 | 0.9112 | 0.9112 |
| (g)Unet++ Resnet50 Adam | 0.4989 | 0.7981 | 0.8849 | 0.8843 | 0.8854 |
| (h )Unet++ Resnet50 RMSprop | 0.5125 | 0.7785 | 0.8715 | 0.8711 | 0.8720 |
| (i)Unet++ ResNeXt50_32x4d Adam | 0.4969 | 0.8214 | 0.9001 | 0.9001 | 0.9000 |
| (k)Unet++ ResNeXt50_32x4d RMSprop | 0.5039 | 0.7139 | 0.8292 | 0.8292 | 0.8292 |

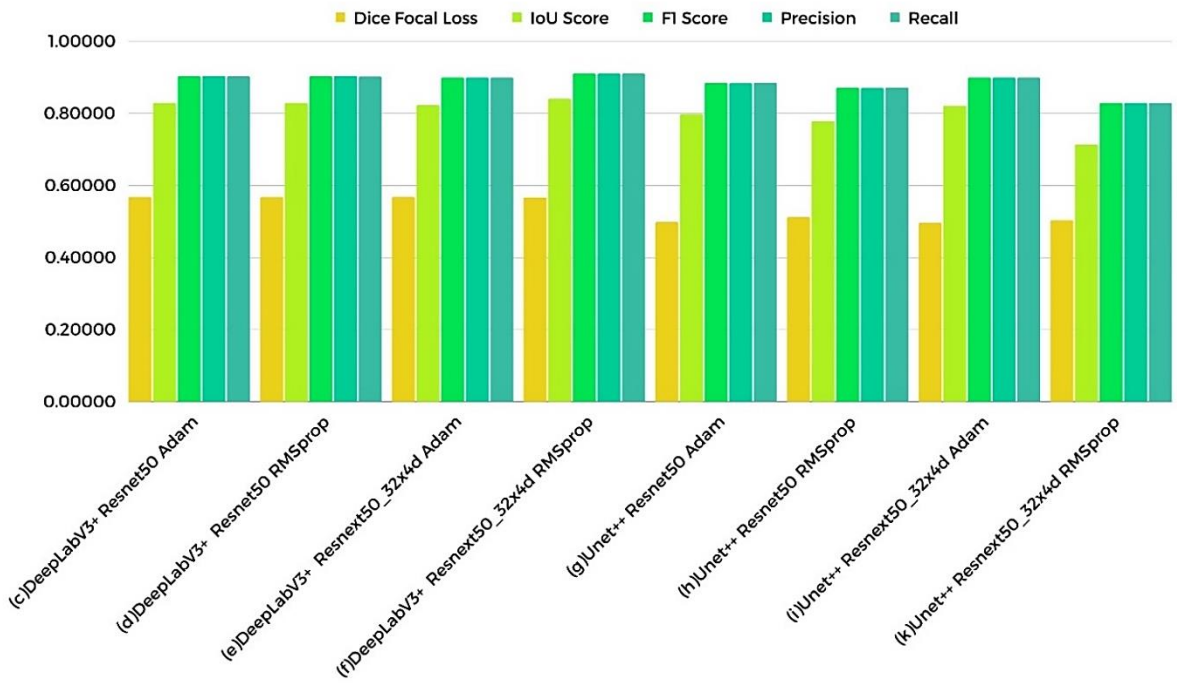


**Figure 7.** Graphical visualization of the evaluation metric scores of the tested models for Dice-Focal loss experiments.

**Table 3.** Metric Evaluations of Tested Models of Focal Loss Experiments

| Experiment / Evaluation Metric | Focal Loss | IOU Score | F-1 Score | Precision | Recall |
|---|---|---|---|---|---|
| (c)DeepLabv3+ Resnet50 Adam | 0.1649 | 0.7935 | 0.8808 | 0.8819 | 0.8797 |
| (d)DeepLabv3+ Resnet50 RMSprop | 0.1634 | 0.8265 | 0.9023 | 0.9034 | 0.9011 |
| (e)DeepLabv3+ ResNeXt50_32x4d Adam | 0.1634 | 0.8261 | 0.9027 | 0.9041 | 0.9013 |
| (f)DeepLabv3+ ResNeXt50_32x4d RMSprop | 0.1640 | 0.8105 | 0.8931 | 0.8942 | 0.8921 |
| (g)Unet++ Resnet50 Adam | 0.1649 | 0.7952 | 0.8810 | 0.8822 | 0.8798 |
| (h )Unet++ Resnet50 RMSprop | 0.1643 | 0.8073 | 0.8893 | 0.8914 | 0.8871 |
| (i)Unet++ ResNeXt50_32x4d Adam | 0.1624 | 0.8524 | 0.9176 | 0.9177 | 0.9175 |
| (k)Unet++ ResNeXt50_32x4d RMSprop | 0.1627 | 0.8442 | 0.9126 | 0.9124 | 0.9128 |

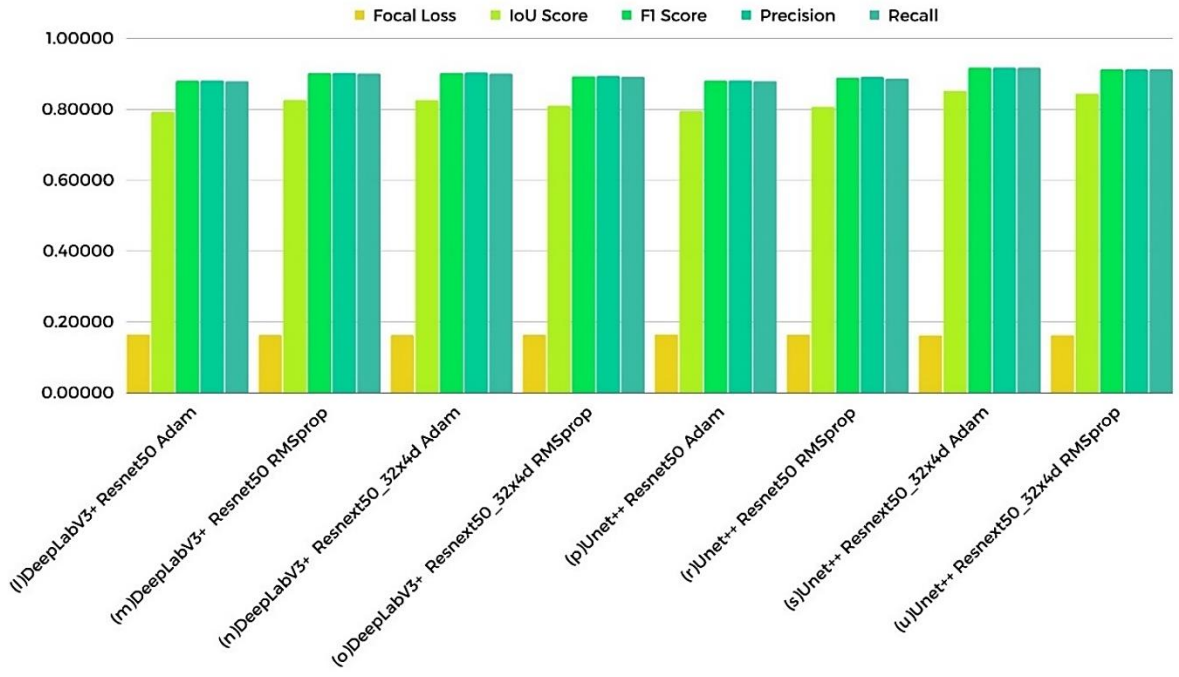


**Figure 8.** Graphical visualization of the evaluation metric scores of the tested models for Focal Loss Experiments.

### 3.2. Visual Evaluation

When Dice-Focal loss and Focal loss are visually compared, it is seen in Figure 9 and Figure 10 that the high one-to-one match rate is higher, especially in forests, inland waters, and agricultural areas. DeepLabv3+ architecture, ResNeXt 50 backbone, and RMSProp for Dice-Focal loss; Unet++ architecture, ResNeXt 50 backbone, and Adam optimizer experiments for Focal loss visually confirmed their higher metrics performance and enabled the creation of LULC maps closest to the real ground mask. The road and rail class and permanent crops class are the most challenging classes to segment, especially for the road and rail class, where located between or under the trees.

In the class-based segmentation evaluation conducted in the experiments, the forest class accounted for 44.1% of the training dataset. Since the training was performed using red, green, blue, and near-infrared (NIR) bands, the detection of the forest class was relatively easier for the models. This is due to its significantly higher reflectance in the green, red, and especially near-infrared bands compared to other classes. Additionally, the forest class did not exhibit significant spectral mixing with other classes, resulting in high F-1 Score. Regarding inland water resources, despite having a low representation of 0.9% in the training dataset, it was classified with minimal error, achieving a very high F-1 Score. Inland water class, unlike other classes, has a high blue reflectance in the visible region and is easier to distinguish based on OBIA than other classes. Although mine-dump and construction sites have a training data set rate of 0.7%, their segmentation in the analysis has been successfully carried out, as can be seen from the high F-1 Score. In the class-based segmentation evaluation conducted in the experiments, the mine-dump and construction class is a distinct land cover type that is relatively easy to distinguish from other classes due to its unique spectral reflection characteristics. It exhibits specific reflectance properties across the red, green, blue, and near-infrared (NIR) bands, making it more identifiable for the models.

The classes most prone to misclassification are found in agricultural areas, where factors such as crop type, harvesting stages, and soil characteristics influence segmentation accuracy. In the training dataset, heterogeneous agricultural lands accounted for 26.5%, arable agricultural lands for 8.9%, and permanent agricultural lands for 5.0%, making up a total of 40.4%.

Among these, permanent and arable agricultural lands were more accurately segmented with minimal mixing into other classes. However, heterogeneous agricultural areas, despite having a larger number of training samples, showed spectral similarities with discontinuous urban texture, as well as road and railway classes, leading to some misclassification. Overall, the models performed well in segmenting agricultural classes, but segmentation of discontinuous urban texture (1% of the training dataset) and road and railway classes (0.5%) did not reach the desired accuracy levels.


**Figure 9**. Predicted masks on test data in Dice-Focal loss experiments.

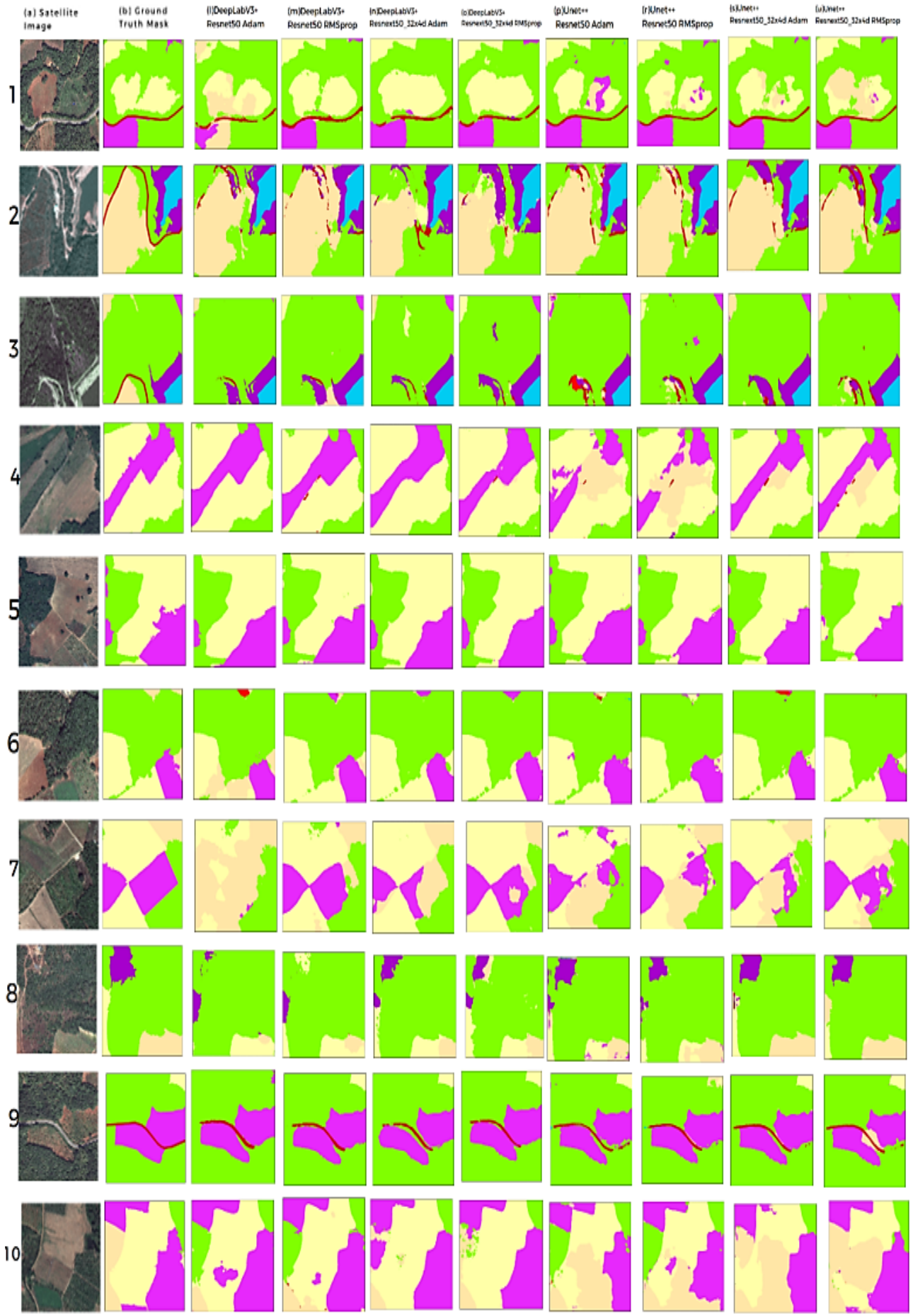


**Figure 10**. Predicted masks on test data in focal loss experiments.

## 4. Discussion

When the LULC maps are evaluated, Unet++ architecture with ResNeXt 50 backbone, Adam optimizer and Focal loss produced the most accurate results with an IoU score of 0.85 and F-1 Score of 0.92. While the experiment of DeepLabv3+ architecture with ResNeXt 50 backbone, RMSProp and Dice-Focal loss is the second most accurate method with an IoU score of 0.84 and F-1 score of 0.91. This shows that the ResNeXt 50 backbone achieves general success in feature extraction, but also shows that different architectures can achieve better compatibility with different optimizers.

As seen in the third row of Figures 9 and 10 and the second row of Figure 11, there are instances where the Very High-Resolution (VHR) WorldView-3 (WV-3) image and the ground truth masks do not perfectly align spatially. This discrepancy suggests that the ground truth masks do not always accurately represent the actual landscape.

For example, in one case, a visible road in the satellite image is labeled as forest in the ground truth mask. Conversely, as seen more clearly in the satellite image, the roads in this region are irregular, branched, and interrupted, making them difficult to detect due to forest coverage. However, in the ground truth mask, these roads appear as if they were detected under ideal conditions (as shown in the second row of Figures 9 and 10 and the first row of Figure 11).

Overall, while the segmentation results generally show a high degree of consistency and harmony, road detection presents a different challenge that should be considered separately from model performance evaluation. Similar issues led to lower F1 scores for irregular and difficult-to-segment classes such as heterogeneous agricultural areas, roads and railways, and urban textures interrupted by roads and railways. These inconsistencies caused deviations between the results and the actual landscape.

Despite these challenges, it can be concluded that LULC maps generated by DL methods provide a more accurate representation of image-based land cover patterns, even if they do not perfectly align with ground truth representations.

This study benefits from the Kaggle platform for optimal performance, cost, and resource efficiency in producing LULC maps. GPUs with higher can produce higher accuracy LULC maps by providing an option to increase epoch number, batch size, and learning rate. The goal is to improve accuracy, effectiveness, time, cost, and resource efficiency by employing DL-based models for continuous analysis and observation.

When we compared DeepLabV3+ and Unet++ architectures for LULC Map production, Unet++ produced more accurate results with Dice-Focal loss. DeepLabV3+, which utilizes Atrous convolution to transfer basic information directly to the decoder, may initially lead to the loss of some fundamental features. However, it effectively distinguishes objects of varying scales, aligning well with the core principles of Object-Based Image Analysis (OBIA). In contrast, UNet++ leverages a skip connection structure to enhance the accurate transfer of low-level features to the decoder. As an OBIA-driven architecture, it has demonstrated strong performance in generating LULC maps, contributing to improved segmentation accuracy.

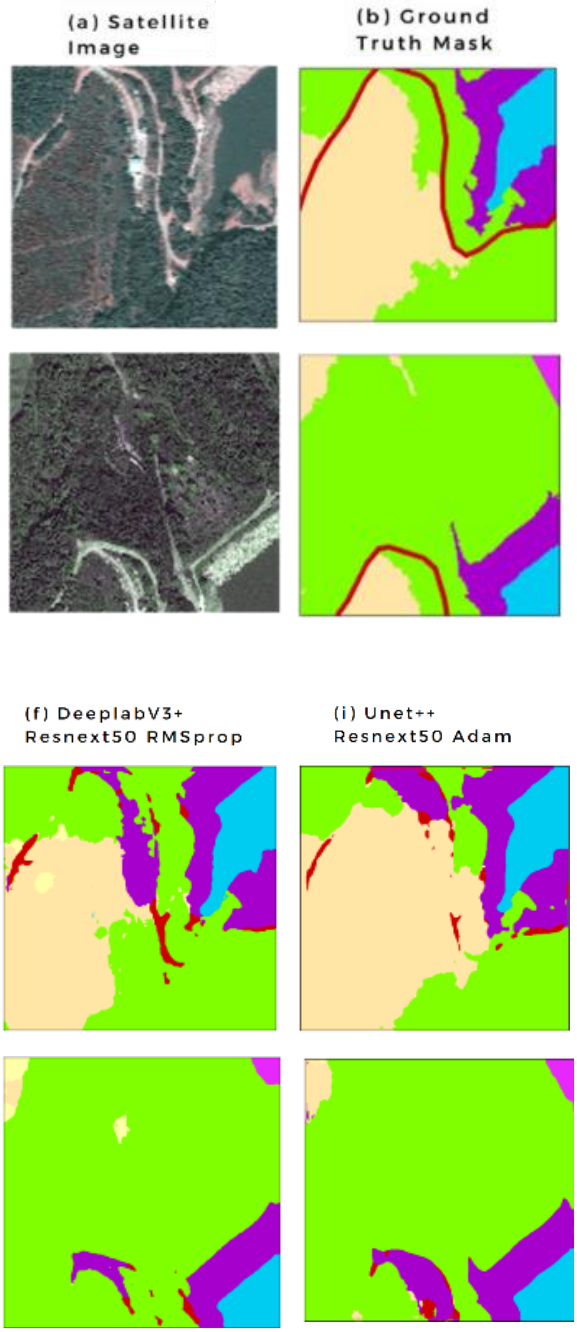


**Figure 11:** a) satellite image, b) ground truth mask, c) the DeeplabV3+ architecture with ResNeXt 50 backbone and RMSProp optimizer, which gives the best result for Dice-Focal loss, and d) the Unet++ architecture with ResNeXt 50 backbone and Adam optimizer, which gives the best result for Focal loss.

ResNet50 is a ResNet model with 50 layers and approximately 23 million parameters. In this study, the ResNet50 backbone was chosen over ResNet101 and ResNet152 due to its lighter architecture, making it more efficient. This decision was influenced by both the constraints of the Kaggle platform and the model's computational effectiveness. Looking at ResNeXt 50, it shares the computational load through group convolutions and feature extraction in parallel and is a backbone with 22 million parameters, which also helps

in high-accuracy segmentation. In this study, ResNeXt 50 backbone gave more effective and higher accuracy results based on different accuracy metrics. This showed that feature extraction with group convolution is more effective in creating LULC maps.

When comparing Adam and RMSprop optimizers, Adam proves to be a more comprehensive and effective choice. While Adam utilizes exponential moving averages of both momentum and RMSprop, the RMSprop optimizer only considers the exponential average of gradient squares. Experimental results indicate that models trained with Adam achieved higher accuracy in key metrics [56,57]. In this context, an experiment utilizing the UNet++ architecture, ResNeXt50 backbone, and Adam optimizer emerged as the most efficient approach for generating a fast, accurate, and optimized LULC map. This suggests that leveraging group convolutions for feature extraction and employing more advanced optimizers such as Adam—alongside modern backbone architectures—can enhance segmentation performance. Additionally, the encoder-decoder structure of UNet++, which enables more precise transmission of fundamental features to the decoder, highlights the importance of feature extraction in LULC mapping. This also explains why UNet++ yielded lower Dice-Focal loss values compared to DeepLabV3+.

A key advantage of the UNet++ + ResNeXt50 + Adam configuration is its ability to successfully segment LULC across all classes, including artificial non-agricultural vegetation areas.

According to Boonpook et al., DL-based semantic segmentation algorithms have proven to be highly effective in detecting and extracting LULC features from images of different spatial resolutions. For example, a study using the LoopNet architecture demonstrated significant segmentation accuracy on medium resolution Landsat 8 imagery, achieving an overall accuracy of 89.84% and a mean intersect of union (mIoU) of 71.69% [58]. Similar to the findings in this study, using advanced architectures such as U-Net++ and DeepLabV3+ for semantic segmentation outperformed pixel-based machine learning methods such as Random Forest and Support Vector Machine, especially in capturing clear boundaries and shapes of LULC classes. Kemker et al. used VHR UAV imagery to identify 18 distinct classes, primarily at the object level [62]. In comparison, their results with an overall accuracy 59.8% was lower than our research. This is a significant finding that supports the idea that the segmentation task gets harder as the complexity and number of terrain classes increase, even though the related study contains highly detailed UAV images. According to Du et al., the accuracy of semantic labeling for VHR images can be significantly increased by combining the OBIA approach with DeepLabv3+. The integration helps alleviate the problem of DeepLabv3+'s inability to handle the spectral heterogeneity of objects, which frequently results in small objects being misclassified. Furthermore, DeepLabv3+ has trouble preserving precise boundary information for ground objects; however, this issue can be resolved when OBIA is used in conjunction with it [59].

In DeepLabv3+ wide-area land cover classification, traditional CNN models such as pixel-based CNNs have difficulty extracting fine spatial details and accurate boundaries for smaller classes. This problem stems from the fact that CNNs typically extract deep spatially relevant features based on fixed input patch sizes that do not effectively account for the varying scales and geometries of objects in heterogeneous landscapes. According to Martins et al., Object-based Convolutional Neural Networks (OCNN), which integrate object segmentation with CNN classification, can address these issues by adjusting the input size and feature extraction process according to object-specific features. Studies have shown that multi-scale CNN architectures, such as those implemented in multi-OCNN frameworks, outperform single CNN models by improving the accuracy of object classification and preserving spatial details, especially for small or complex objects [60]. Zhang et al. identified five distinct LULC classes and one additional class using the Atrous spatial pyramid pooling (ASPP)-UNet model. They used Beijing city's WorldView-2 (WV-2) and WV-3 images to train and evaluate their suggested model. They achieved an overall accuracy of 84.0% for six classes of WV-3 test images [37]. Since Sertel et al.'s study, to which the dataset used in this study belongs, and also Zhang et al.'s study used similar VHR images, the class-wise accuracy in common classes was also examined. While Sertel et al. obtained F-1 Scores of 0.612 and 0.983 for the Aksu test dataset, Zhang et al. obtained F-1 Scores of 0.906 and 0.755 for the water and road classes, respectively. The F-1 Scores of the related Aksu dataset are supported by this study [13,37].

Using Sentinel-2, WV-2, and Pleiades-1B satellite imagery, Bengana et al. employed six distinct LULC classes within the classes identified in this study, as well as a generalized CORINE Land Cover nomenclature as ground truth. They combined various urban density classes, such as classes related to industry and mining, into a single class called urban for the purposes of their study. Additionally, they unified all agricultural classes—including permanent crops, arable land, and diverse regions—under a single agricultural class. They obtained an average IoU value of 0.5559 for six classes with WV-2 imagery, while we obtained an average IoU value of 0.8524 for eight LULC classes in this study [61].

According to Sertel et al. [13], the DeepLabV3+ architecture combined with the ResNeXt50 backbone achieved the highest accuracy in generating LULC maps. While the related study used 12 CORINE classes, this study focuses on 8 CORINE classes instead. Additionally, the learning rate in this study is set to $10^{-5}$ instead of $10^{-4}$. Beyond the previous study, this research also incorporates Focal loss and evaluates performance using both the RMSprop and Adam optimizers. The results confirm the findings of the related study, as DeepLabV3+ with Dice-Focal loss, ResNeXt50 backbone, and RMSprop optimizer achieved high accuracy. Furthermore, the UNet++ architecture with Focal loss and Adam optimizer also produced highly accurate LULC maps when paired with ResNeXt50 and the Adam optimizer. These findings highlight the impact of the number of classes on LULC map accuracy when using ResNeXt50 and the Adam

optimizer, underscoring the importance of dataset composition. Additionally, it was observed that the effectiveness of loss functions varies across different architectures, further emphasizing the influence of model design on segmentation performance.

## 5. Conclusion

LULC mapping from VHR satellite images is an effective method in terms of accuracy, time, cost, and resources. However, performing the segmentation of individual VHR satellite images, evaluating their accuracy, and producing maps based on this is a comprehensive work that requires extensive time and human resources. In this context, CNN-based segmentation is of great importance, especially to automatically generate LULC maps and identify long-term time-dependent changes in different geographical regions.

This study focused on the segmentation of the Aksu region, where the primary land cover types in the image dataset consist of 44.1% forest and 40.4% agricultural land. This study region and other regions with similar landscape characteristics reveal the importance of LULC maps, especially to better model and determine fires, natural disasters, and urbanization activities, which threaten forests and agricultural areas.

In this study, the impact of class distributions on segmentation accuracy was examined, along with the effects of loss functions, backbones, batch size, learning rate, and iteration size. According to the findings, in Focal loss experiments, the UNet++ architecture, ResNeXt50 backbone, and RMSprop optimizer produced the most accurate and high-performing LULC maps. This was achieved with a learning rate of $10^{-5}$, a batch size of 4, and training over 100 epochs, yielding superior results in IoU, precision, recall, and Dice-Focal loss metrics.

The results further demonstrate that, regardless of the ground truth labels, errors in segmentation caused by human factors can be mitigated in LULC mapping through the proposed methods.

These findings emphasize the potential of DL-based segmentation approaches in improving the accuracy and reliability of LULC segmentation. Future research could explore the integration of multi-temporal satellite imagery and additional spectral bands, enabling more robust monitoring of landscape changes over time. Furthermore, adapting these models to different geographic regions and environmental conditions could enhance their generalizability, providing a scalable and automated solution for large-scale LULC mapping.

## Author contributions

**İskender Berkay Sür:** Conceptualization, Methodology, Writing-Original draft. **Ugur Alganci:** Writing-Original draft preparation, Methodology, Validation. **Elif Sertel:** Visualization, Methodology, Investigation, Supervision, Writing-Reviewing and Editing.

## Conflicts of interest

The authors declare no conflicts of interest.

## References


1. Treitz, P., & Rogan, J. (2004). Remote sensing for mapping and monitoring land-cover and land-use change-an introduction. Progress in Planning, 61, 269-279.https://doi.org/10.1016/S0305-9006(03)00064-3
2. Mora, B., Tsendbazar, N., Herold, M., & Arino, O. (2014). Global Land Cover Mapping: Current Status and Future Trends. In: Manakos, I., Braun, M. (eds) Land Use and Land Cover Mapping in Europe. Remote Sensing and Digital Image Processing, vol 18. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-7969-3_2
3. Rogan, J., & Chen, D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. Progress in Planning, 61, 301-325. https://doi.org/10.1016/S0305-9006(03)00066-7.
4. Saleem, A., & Mahmood, S. (2023). Spatio-temporal assessment of urban growth using multi-stage satellite imageries in Faisalabad, Pakistan. Advanced Remote Sensing, 3(1), 10–18.
5. Zadbagher, E., Marangoz, A. M., & Becek, K. (2023). Characterizing and estimating forest structure using active remote sensing: An overview. Advanced Remote Sensing, 3(1), 38–46.
6. Efe, E., & Alganci, U. (2023). Çok zamanlı Sentinel 2 uydu görüntüleri ve makine öğrenmesi tabanlı algoritmalar ile arazi örtüsü değişiminin belirlenmesi. Geomatik, 8(1), 27-34. https://doi.org/10.29128/geomatik.1092838
7. Pala, İ., & Alganci, U. (2025). Investigating the performance of super-resolved remote sensing images on coastline segmentation with deep learning based methods. International Journal of Engineering and Geosciences, 10(1), 93-106. https://doi.org/10.26833/ijeg.1522143
8. Carter, S., & Herold, M. (2019). Specifications of land cover datasets for SDG indicator monitoring. Retrieved from: https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification-for-globalindicators/10th-meeting---september-2019/10.2_--UNEP-WCMC_Sarah_Carter.pdf.
9. Nguyen, H. Q., Doan, T. D., Tomppo, E., & McRoberts, R. E. (2020). Land Use/Land Cover Mapping Using Multitemporal Sentinel-2 Imagery and Four Classification Methods—A Case Study from Dak Nong, Vietnam. Remote Sensing, 12(9), 1367. https://doi.org/10.3390/rs12091367.
10. Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2217. https://doi.org/10.1109/JSTARS.2019.2918242.
11. Shi, Y., Qi, Z., Liu, X., Niu, N., Zhang, H. (2019). Urban land use and land cover classification using multisource remote sensing images and social media data. Remote Sensing, 11, 2719. https://doi.org/10.3390/rs11222719.

12. Uba, N. K. (2019). Land Use and Land Cover Classification Using Deep Learning Techniques. arXiv:1905.00510.https://doi.org/10.48550/arXiv.1905.00510.
13. Sertel, E., Ekim, B., Osgouei, P. E., & Kabadayi, M. E. (2022). Land Use and Land Cover Mapping Using Deep Learning Based Segmentation Approaches and VHR Worldview-3 Images. Remote Sensing, 14(18), 4558. https://doi.org/10.3390/rs14184558.
14. Henry, C. D., Storie, C. D., Palaniappan, M., Alhassan, V., Swamy, M., Aleshinloye, D. K., Curtis, A., & Kim, D. (2019). Automated LULC map production using deep neural networks. International Journal of Remote Sensing, 40(11), 4416–4440. https://doi.org/10.1080/01431161.2018.1563840
15. Long, J., Shelhamer, E., Darrell, T., (2015). Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition. 07-12-June-2015, 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965.
16. Bakırman, T., & Sertel, E. (2023). A benchmark dataset for deep learning-based airplane detection: HRPlanes. International Journal of Engineering and Geosciences, 8(3), 212-223. https://doi.org/10.26833/ijeg.1107890.
17. Akar, Ö., Saralıoğlu, E., Güngör, O., Bayata, H. F. (2024). Semantic segmentation of very-high spatial resolution satellite images: A comparative analysis of 3D-CNN and traditional machine learning algorithms for automatic vineyard detection. International Journal of Engineering and Geosciences, 9(1), 12-24. https://doi.org/10.26833/ijeg.1252298.
18. Nasiri, V., Deljouei, A., Moradi, F., Sadeghi, S. M. M., & Borz, S. A. (2022). Land Use and Land Cover Mapping Using Sentinel-2, Landsat-8 Satellite Images, and Google Earth Engine: A Comparison of Two Composition Methods. Remote Sensing, 14(9), 1977. https://doi.org/10.3390/rs14091977.
19. Sertel, E., Kabadayı, M.E., Sengul, G.S., & Tumer, I.N. (2024). HexaLCSeg: A historical benchmark dataset from Hexagon satellite images for land cover segmentation [Software and Data Sets]. IEEE Geoscience and Remote Sensing Magazine, 12, 197-206. https://doi.org/10.1109/MGRS.2024.3394248.
20. Zhang, Y. (2002). A new automatic approach for effectively fusing Landsat 7 as well as IKONOS images. In IEEE International Geoscience and Remote Sensing Symposium, 4, 2429-2431. http://dx.doi.org/10.1109/IGARSS.2002.1026567.
21. Wang, P., & Sertel, E. (2021). Channel–spatial attention-based pan-sharpening of very high-resolution satellite images. Knowledge-Based Systems, 229, 107324. https://doi.org/10.1016/j.knosys.2021.107324.
22. Sklearn Package. (2023) Sklearn. Last Access: 30 December 2023, https://scikit-learn.org/stable/about.html#citing-scikit-learn.
23. Ekim, B., Sertel, E., Kabadayı, M.E. (2021). Automatic Road extraction from historical maps using deep learning techniques: A regional case study of Turkey in a German World War II Map. ISPRS International Journal of Geo-Information, 10(8), 492. https://doi.org/10.3390/ijgi10080492.
24. Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. Journal of the Royal Statistical Society, Series B (Methodological), 36(2), 111–147.
25. Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning. In Springer series in statistics.https://doi.org/10.1007/978-0-387-84858-7.
26. Joseph, V. R., & Vakayil, A. (2021). SPLIT: an optimal method for data splitting. Technometrics, 64(2), 166–176. https://doi.org/10.1080/00401706.2021.1921037.
27. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J. T., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Liu, F., & Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. arXiv:1912.01703 https://doi.org/10.48550/arxiv.1912.01703.
28. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, 248-255. https://doi.org/10.1109/CVPR.2009.5206848.
29. Mandt, S., Hoffman, M. D., & Blei, D. M. (2017). Stochastic gradient descent as approximate bayesian inference. Journal of Machine Learning Research, 18(134), 1-35. https://doi.org/10.48550/arXiv.1704.04289.
30. Smith, S. L., Duckworth, D., Rezchikov, S., Le, Q. V., & Sohl-Dickstein, J. (2018). Stochastic natural gradient descent draws posterior samples in function space. arXiv:1806.09597 https://doi.org/10.48550/arXiv.1806.09597.
31. Kaggle. (2023). Kaggle: your machine learning and data science community. https://www.kaggle.com/.
32. Castelluccio, M., Poggi, G., Sansone, C., & Verdoliva, L. (2015). Land use classification in remote sensing images by convolutional neural networks. arXiv.1508.00092. https://doi.org/10.48550/arXiv.1508.00092.
33. Luus, F. P., Salmon, B. P., van den Bergh, F., & Maharaj, B. (2015). Multiview deep learning for land use classification. IEEE Geoscience and Remote Sensing Letters, 12(12), 2448-2452. https://doi.org/10.1109/LGRS.2015.2483680.
34. Marmanis, D., Datcu, M., Esch, T., & Stilla, U. (2016). Deep learning earth observation classification using ImageNet pretrained networks. IEEE Geoscience and Remote Sensing Letters, 13(1), 105-109. https://doi.org/10.1109/LGRS.2015.2499239.
35. Ünel, F. B., Kuşak, L., Yakar, M., & Doğan, H. (2023). Coğrafi bilgi sistemleri ve analitik hiyerarşi prosesi

kullanarak Mersin ilinde otomatik meteoroloji gözlem istasyonu yer seçimi. *Geomatik*, *8*(2), 107-123.

36. Alhassan, V., Henry, C. D., Ramanna, S., & Storie, C. D. (2019). A deep learning framework for land-use/land-cover mapping and analysis using multispectral satellite imagery. Neural Computing and Applications, 32(12), 8529–8544. https://doi.org/10.1007/s00521-019-04349-9.
37. Zhang, P., Ke, Y., Zhang, Z., Wang, M., Li, P., & Zhang, S. (2018). Urban Land Use and Land Cover Classification Using Novel Deep Learning Models Based on High Spatial Resolution Satellite Imagery. Sensors, 18(11), 3717. https://doi.org/10.3390/s18113717.
38. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834-848. https://doi.org/10.1109/TPAMI.2017.2699184.
39. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV), 801-818. https://doi.org/10.48550/arXiv.1802.02611.
40. Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2019). Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging, 39(6), 1856-1867. https://doi.org/10.1109/TMI.2019.2959609.
41. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778. https://doi.org/10.1109/CVPR.2016.90.
42. He, K., & Sun, J. (2015). Convolutional neural networks at constrained time cost. In Proceedings of the IEEE conference on computer vision and pattern recognition, 5353-5360. https://doi.org/10.48550/arXiv.1412.1710.
43. Srivastava, R. K., Greff, K., & Schmidhuber, J. (2015). Highway networks. https://doi.org/10.48550/arXiv.1505.00387.
44. Stoyanov, D., Taylor, Z., Carneiro, G., Syeda-Mahmood, T., Martel, A., Maier-Hein, L., ... & Madabhushi, A. (Eds.). (2018). Deep learning in medical image analysis and multimodal learning for clinical decision support. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: : 4th International Workshop DLMIA 2018 and 8th International Workshop ML-CDS 2018 Held in Conjunction With MICCAI 2018 Granada Spain September 20 2018 Proceedings, Cham, Switzerland:Springer, vol. 11045, 2018.
45. Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision, 2980-2988. https://doi.org/10.1109/ICCV.2017.324.
46. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M. (2017). Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In: Cardoso, M., et al. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA ML-CDS 2017 2017. Lecture Notes in Computer Science (), vol 10553. Springer, Cham. http://doi.org/10.1007/978-3-319-67558-9_28.
47. Mulyanto, M., Faisal, M., Prakosa, S. W., & Leu, J. S. (2021). Effectiveness of focal loss for minority classification in network intrusion detection systems. Symmetry, 13(1), 4. http://doi.org/10.3390/sym13010004.
48. Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. Journal of Machine Learning Research, 12(7), 2121-2159.
49. Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. arXiv:1212.5701. https://doi.org/10.48550/arXiv.1212.5701
50. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv:1412.6980 https://doi.org/10.48550/arXiv.1412.6980.
51. Teferi, E., Bewket, W., Uhlenbrook, S., & Wenninger, J. (2013). Understanding recent land use and land cover dynamics in the source region of the Upper Blue Nile, Ethiopia: Spatially explicit statistical modeling of systematic transitions. Agriculture, Ecosystems and Environment, 165, 98-117. https://doi.org/10.1016/j.agee.2012.11.007.
52. Powers, D.M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv:2010.16061. https://doi.org/10.48550/arXiv.2010.16061.
53. Congalton, R. G., & Green, K. (2008). Assessing the Accuracy of Remotely Sensed Data. In CRC Press eBooks. https://doi.org/10.1201/9781420055139
54. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 658-666. https://doi.org/10.48550/arXiv.1902.09630.
55. Längkvist, M., Kiselev, A., Alirezaie, M., Loutfi, A. (2016). Classification and segmentation of satellite orthoimagery using convolutional neural networks. Remote Sensing, 8, 329. http://doi.org/10.3390/rs8040329
56. Cheng, G., Xie, X., Han, J., Guo, L., Xia, G.-S. (2020). Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 3735–3756. http://doi.org/10.1109/JSTARS.2020.3005403
57. Yuan, X., Shi, J., Gu, L. (2021). A Review of Deep Learning Methods for Semantic Segmentation of Remote Sensing Imagery. Expert Systems with Applications, 169, 114417.

http://doi.org/10.1016/j.eswa.2020.114417
58. Boonpook, W., Tan, Y., Nardkulpat, A., Torsri, K., Torteeka, P., Kamsing, P., ... & Jainaen, M. (2023). Deep learning semantic segmentation for land use and land cover types using Landsat 8 imagery. ISPRS International Journal of Geo-Information, 12(1), 14. https://doi.org/10.3390/ijgi12010014
59. Du, S., Du, S., Liu, B., Zhang, X. (2020). Incorporating DeepLabv3+ and object-based image analysis for semantic segmentation of very high resolution remote sensing images. International Journal of Digital Earth, 14, 1–22. http://doi.org/10.1080/17538947.2020.1831087
60. Martins, V. S., Kaleita, A. L., Gelder, B. K., da Silveira, H. L., & Abe, C. A. (2020). Exploring multiscale object-based convolutional neural network (multi-OCNN) for remote sensing image classification at high spatial resolution. ISPRS Journal of Photogrammetry and Remote Sensing, 168, 56-73.
61. Bengana, N., Heikkila, J. (2021). Improving land cover segmentation across satellites using domain adaptation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 1399–1410. http://doi.org/10.1109/JSTARS.2020.3042887
62. Kemker, R., Salvaggio, C., Kanan, C. (2018). Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. ISPRS Journal of Photogrammetry and Remote Sensing, 145, 60–77. http://doi.org/10.1016/j.isprsjprs.2018.04.014