Optimized Deep Learning Model for LULC Classification in the Terai Region: Comparing U-Net and Resunet with Spectral Index Fusion and Cross-Area Fine-Tuning

Hiralal Pardhe[1], Uma Shankar Panday[2] & Sanjeevan Shrestha[3]
hiralalpardhe@gmail.com, uspanday@ku.edu.np, shr.sanjeevan@gmail.com
[1]Survey Office Kalanki, [2]Department of Geomatics Engineering, Kathmandu University,
[3]Ministry of Land Management, Co-operatives, Federal Affairs and General Administration

KEYWORDS

Deep learning, U-Net, ResNet-50, Cross-area fine-tuning, Hyperparameter optimization, Spectral indices, GIS, Environmental monitoring

ABSTRACT

Accurate and up-to-date Land Use/Land Cover (LULC) mapping is crucial for environmental management and regional planning, but its development is often limited by insufficient labeled data, spectral similarity between land-cover classes, and difficulties in transferring models across regions. This study developed an optimized deep learning framework for LULC classification in Dhanusha District using cross-area fine-tuning with U-Net and a ResUNet-style architecture based on a ResNet-50 encoder backbone. The approach aimed to improve model transferability by adapting models trained in a data-rich source area (Mahottari District) to a target area (Dhanusha District) with limited labeled samples. Sentinel-2 Level-2A imagery from March 2021 was preprocessed and five LULC classes - Water Bodies, Cropland, Built-Up, Tree Cover, and Others were manually digitized from Google Earth imagery and converted into raster masks. Along with RGB imagery, spectral-index composites including NDVI, SAVI, EVI, MSAVI, GNDVI, NDWI, and NDBI were generated to improve class separability. Image and mask datasets were tiled into sizes of 128×128, 256×256, and 512×512 pixels and divided into training and testing sets. Hyperparameters such as batch size, dropout, L2 regularization, learning-rate scheduling, and early stopping were systematically optimized. Results demonstrated strong segmentation performance after cross-area adaptation. ResNet-50 consistently outperformed U-Net, achieving overall accuracies of up to 0.96 in Mahottari and 0.95 in Dhanusha, while U-Net achieved up to 0.95 and 0.94, respectively. Spectral-index composites reduced confusion among challenging classes, especially water bodies and built-up areas. Intermediate tile sizes (256×256) provided the best balance between spatial context, computational efficiency, and stable model convergence, resulting in improved generalization across districts.

1. INTRODUCTION

1.1 Background

Land Use/Land Cover (LULC) classification is a fundamental aspect of environmental monitoring, urban planning, and natural resource management. It involves categorizing Earth's surface into distinct classes such as forests, agricultural land, urban areas, water bodies, and barren land (Turner et al., 2007). Accurate LULC mapping is crucial for understanding ecological changes, assessing climate impacts, and formulating sustainable land management policies. Remote sensing (RS) and Geographic Information Systems (GIS) have revolutionized LULC studies by enabling large-scale, cost-effective monitoring through satellite imagery such as Landsat and Sentinel-2 (Drusch et al., 2012).

Nepal has experienced significant LULC changes due to rapid urbanization, agricultural expansion, and deforestation (Uddin et al., 2015).The Terai region, including Dhanusha District, has seen dramatic transformations, with forests converted to farmland (Paudel et al., 2018). These changes have ecological consequences, including biodiversity loss, soil degradation, and increased flood risks. Monitoring these trends is essential for sustainable development, yet traditional field surveys are labor-intensive and have limited scope.

Optical remote sensing has been widely used for LULC studies, leveraging platforms like Landsat (30m resolution) and Sentinel-2 (10m/20m resolution). Early studies employed pixel-based classification methods such as Maximum Likelihood and ISODATA (Richards & Jia, 1999), but these struggled with spectral complexity in heterogeneous landscapes. Machine learning (ML) techniques, including Random Forest (RF) and Support Vector Machines (SVM), improved accuracy by handling non-linear data relationships (Mountrakis et al., 2011). However, deep learning (DL) approaches, particularly Convolutional Neural Networks (CNNs), have recently outperformed traditional ML methods in LULC classification (Zhang et al., 2019).

Earlier approaches primarily relied on pixel-based classification methods such as Maximum Likelihood, which often struggled in heterogeneous landscapes (Richards & Jia, 1999). More recent studies have adopted machine learning algorithms such as Random Forest and Support Vector Machines to improve classification accuracy; however, these methods still depend heavily on handcrafted features and large labeled datasets (Mountrakis et al., 2011).

Despite these advancements, several limitations remain. Many Nepal-based studies have not fully leveraged deep learning techniques capable of capturing spatial context, and the scarcity of labeled data continues to hinder model performance. Furthermore, most existing research focuses on single-region analysis, with limited attention to model transferability across geographically similar areas. This study addresses these gaps by employing cross-area fine-tuning to transfer learned representations between similar regions, thereby reducing dependence on extensive training data while improving generalization (Tuia et al., 2016; Yuan et al., 2020). Additionally, the integration of hyperparameter optimization enhances model efficiency and predictive performance (Akiba et al., 2019). Together, these approaches provide a robust and scalable framework for accurate LULC classification in data-constrained environments like Nepal.

Among deep learning architectures, U-Net (Ronneberger et al., 2015) has emerged as a leading model for semantic segmentation in remote sensing. Its encoder-decoder structure with skip connections preserves spatial details, making it ideal for precise land cover mapping. ResUNet, an extension of

U-Net with residual blocks (Diakogiannis et al., 2020), further enhances feature learning by mitigating vanishing gradients in deep networks. These models have demonstrated superior performance in complex landscapes, achieving over 90% accuracy in LULC classifications (Maggiori et al., 2017).

In this study, a model initially trained on the Mahottari district, which shares comparable agro-ecological and climatic characteristics, is fine-tuned for Dhanusha district. This approach significantly reduces the dependency on extensive ground-truth data while preserving classification accuracy. Previous studies have demonstrated that spatially informed fine-tuning strategies substantially improve LULC mapping performance in data-scarce regions, especially when source and target areas exhibit similar land-use patterns and phenological behavior (Zhu et al., 2017;Gevaert & García-Haro, 2015).

Complementing cross-area fine-tuning, hyperparameter optimization (HPO) plays a critical role in maximizing model efficiency and predictive performance. Key hyperparameters-such as learning rate, batch size, and network depth, strongly influence convergence speed and classification accuracy (Bergstra & Bengio, 2012). Manual tuning is often inefficient and suboptimal; therefore, automated HPO techniques such as Bayesian Optimization and evolutionary search methods, implemented through open-source frameworks like Optuna, allow systematic and computationally efficient exploration of the hyperparameter space (Akiba et al., 2019). The combined use of cross-area fine-tuning and HPO forms a robust framework that enhances model adaptability, accuracy, and computational efficiency for LULC classification in data-constrained remote sensing applications. By utilizing freely available Sentinel-2 imagery and its corresponding composites of spectral indices and Deep learning architectures, this research provides a cost-effective solution for land cover mapping in Nepal.

1.2 Objectives

The primary objective of this study is to develop an **optimized cross-area fine-tuning model** for Land Use/Land Cover (LULC) classification in the **Dhanusha district** using **U-Net and ResUNet** architectures enhanced with **hyperparameter optimization**. The secondary objectives are as follows:

- To access the baseline classification accuracy of U-Net and ResNet-50 models trained on the Mahottari district prior to cross-area fine-tuning.
- To measure the accuracy loss of both models after fine-tuning on the spatially distinct Dhanusha dataset.

2. MATERIALS AND METHODS

2.1 Study area

The study was conducted in the neighboring districts of Mahottari and Dhanusha in Madhesh Province, southeastern Nepal, located in the Terai region. These districts were selected to evaluate the cross-area fine-tuning capability of deep learning–based Land Use Land Cover (LULC) classification models, with Mahottari used as the training area and Dhanusha as the testing area. Their similar climate, topography, and land-use patterns make them suitable for assessing model transferability and generalization.

Mahottari District covers about 1,002 km² and consists mainly of flat Terai plains with diverse land-cover features such as agricultural land, forests, settlements, water bodies, and built-up areas, making it suitable for model training. Dhanusha District, covering approximately 1,180 km², shares similar environmental and land-use characteristics with Mahottari, making it an appropriate region for testing the performance of the transferred deep learning models.

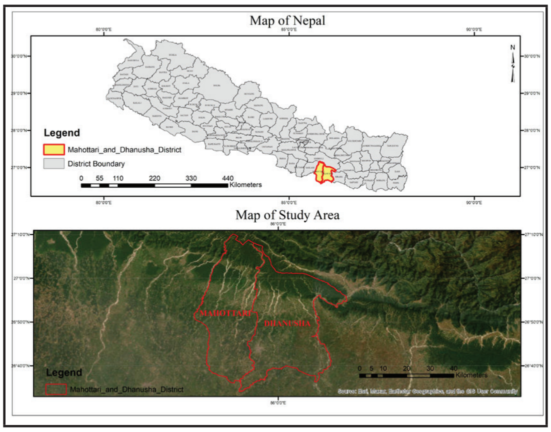


Figure 1: Study area

2.2 Methodology

The overall methodology of the study is presented in *Figure 2*.

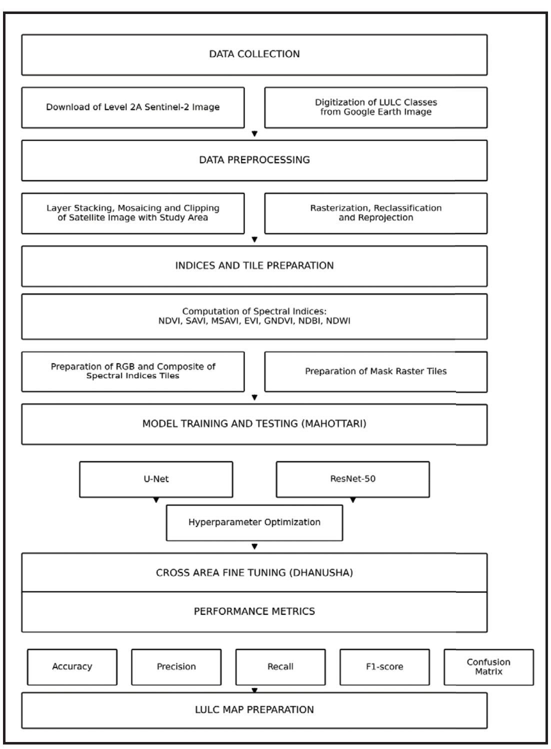


Figure 2: Overall methodology of the study

2.2.1 Data collection

The data acquisition process for Sentinel-2A imagery of March 2021 for Mahottari and Dhanusha districts involves the collection of high-resolution multispectral satellite images captured by the European Space Agency's (ESA) Copernicus Sentinel-2 mission. These images, available at a 10-meter spatial resolution, were downloaded from the Copernicus Open Access Hub. The selection of the March 2021-time frame ensures minimal cloud cover and optimal visibility conditions for surface feature extraction during the dry season. In this study, Sentinel-2 satellite imageries of the year 2021 March of the study area were used. For Mahottari, two imageries tiles (T45RUK and T45RVK) were downloaded. Similarly, three S2 imageries tiles (T45RUK, T45RUL and T45RVK) were downloaded for Dhanusha district. The acquired imagery includes multiple spectral bands, particularly those in the visible, near-infrared (NIR), and short-wave infrared (SWIR) ranges, which are essential for computing spectral indices such as NDVI (Normalized Difference Vegetation Index), NDWI (Normalized Difference Water Index), and NDBI (Normalized Difference Built-up Index). The systematic acquisition of these images serves as the foundational step for further analysis and the development of deep learning-based classification models for accurate LULC mapping.

The labelled dataset for LULC classification in Mahottari and Dhanusha districts was prepared primarily through manual digitization of high-resolution Google Earth imagery from the year 2021. High-resolution Google Earth imagery was visually interpreted to delineate detailed land cover features such as built-up areas, cropped areas, water bodies, tree cover, and other relevant classes. The number of samples for each class is presented in *Table 1*.

Table 1:Classwise Samples of both districts

Class	No of Samples			
	Mahottari		Dhanusha	
	Train	Test	Train	Test
Water Bodies	276	69	152	38
Cropland	1680	420	960	240
Built-Up	1760	440	952	238
Tree Cover	1200	300	720	180
Others	880	220	480	120

2.2.2 Pre-processing

Data preprocessing is a critical stage in the methodology, as it ensures the quality, consistency, and spatial compatibility of both satellite imagery and reference data prior to deep learning-based analysis. All preprocessing operations were carefully performed to minimize noise, reduce geometric discrepancies, and prepare datasets for accurate pixel-wise classification. Initially, the downloaded Sentinel-2 Level-2A multispectral imageries are processed by selecting the relevant spectral bands required for LULC classification.

In parallel, the reference LULC data obtained through manual digitization from Google Earth imagery are prepared for model training. Since deep learning-based semantic segmentation requires pixel-level labels, the vector-based LULC polygons are converted into raster format through rasterization. During rasterization, each pixel is assigned a class label corresponding to the underlying LULC category. The rasterized LULC data are then subjected to reclassification, where original class codes are standardized and harmonized to maintain consistency across the dataset and avoid ambiguities during training.

2.2.2.1 Computation of spectral indices

To enhance feature discrimination among different land cover classes, various **spectral indices** are computed from the preprocessed Sentinel-2 bands. These indices include **NDVI, SAVI, MSAVI, EVI, GNDVI, NDBI, and NDWI**, each designed to emphasize specific surface characteristics such as vegetation vigor, soil influence, built-up areas, and water bodies. The inclusion of these indices improves the deep learning model's ability to distinguish between spectrally similar classes. The derived indices were stacked together to form multi-band composite images, providing comprehensive spectral information for subsequent model training and classification. This composite dataset served as the input for both traditional and deep learning algorithms in the LULC classification process. *Table 2* shows the spectral indices used in the study.

2.2.2.2 Generation of image and mask tiles

Following the steps mentioned above, all three generated tile sizes for the year 2021 were split into training and testing subsets. Twenty percent of the total dataset was kept for validating the performance of the model, and the rest was used for the training of the model. Similarly, the tiles corresponding to the Dhanusha district were separated in the same manner to facilitate fine-tuning and testing of performances of models.

2.2.3 Model training and testing

Model training and testing were performed using labeled samples from the Mahottari district, which served as the source area for subsequent cross-area fine-tuning. The available dataset was divided into training and testing subsets to ensure reliable model development and unbiased performance assessment. Training was carried out iteratively by optimizing model parameters to minimize classification error, with training progress monitored to ensure stable convergence. To enhance generalization under limited data conditions, appropriate regularization strategies and a controlled learning rate were applied during training.

2.2.3.1 U-net

The implemented model is based on the **U-Net architecture**, a fully convolutional neural network (FCN) that has proven highly effective for **semantic segmentation tasks.** U-Net is specifically designed to identify and classify each pixel into distinct land-cover categories such as, water, built-up areas, cropland, and tree cover. The architecture follows a **symmetric encoder - decoder structure** with skip connections that bridge corresponding layers between the encoder and decoder, ensuring that detailed spatial information is preserved while the network learns broader contextual features.

The **encoder (contracting path)** extracts increasingly abstract spatial spectral features from the input images through successive convolutional blocks, with filter depths of 32, 64, 128, and 256. Each block is composed of two convolutional layers followed by batch normalization, **ReLU activation**, **dropout**, and **max pooling** for downsampling. The **dropout rate** was optimized between **0.05 and 0.1** to balance regularization and feature learning, preventing overfitting without hindering model convergence. The **bottleneck layer**, consisting of 512 filters, captures deep and compact feature representations of the image.

The **decoder (expansive path)** restores the spatial dimensions using **transposed convolutions (up-convolutions)** and concatenates the corresponding encoder outputs through skip connections, enabling the network to reconstruct high-resolution segmentation maps with accurate boundary delineation. The final 1*1 convolution layer with a **softmax activation function** generates a five-channel output, representing the probability of each pixel belonging to one of the land-cover classes (Ronneberger et al., 2015).

Table 2: Spectral indices used in the study

Indices	Full Name	Formula	Purpose	References
NDVI	Normalized Difference Vegetation Index	$\frac{\text{NIR-RED}}{\text{NIR+RED}}$	Separates vegetation from non- vegetation classes	(Rouse et al., 1973)
SAVI	Soil Adjusted Vegetation Index	$\frac{(NIR - Red)}{(NIR + Red + L)} \times (1 + L)$, where $L = 0.5$	Improves vegetation detection in sparse or soil-exposed areas	(Huete, 1988)
EVI	Enhanced Vegetation Index	$2.5 \times \frac{(NIR - Red)}{(NIR + 6 \times Red - 7.5 \times Blue + 1)}$	Enhances dense vegetation mapping and reduces atmospheric noise	(Huete et al., 2002)
MSAVI	Modified Soil Adjusted Vegetation Index	$\frac{2 \times NIR + 1 - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - Red)}}{2}$	Distinguishes vegetation in mixed soil-vegetation pixels	(Qi et al., 1994)
GNDVI	Green Normalized Difference Vegetation Index	$\frac{\text{NIR-GREEN}}{\text{NIR+GREEN}}$	Detects vegetation health variations and crop conditions	(Gitelson et al., 1996)
NDWI	Normalized Difference Water Index	$\frac{\text{GREEN-NIR}}{\text{GREEN+NIR}}$	Extracts water bodies and separates water from land	(McFEETERS, 1996)
NDBI	Normalized Difference Built-Up Index	$\frac{\text{SWIR-NIR}}{\text{SWIR+NIR}}$	Identifies built-up/urban areas from vegetation and water	(Zha et al., 2003)

2.2.3.2 ResNet-50

The **ResNet-50 architecture** was conceptually integrated into the study as a potential **encoder backbone** for enhanced feature extraction in future work. **ResNet-50 (Residual Network with 50 layers)** introduces the principle of **residual learning**, where shortcut (identity) connections allow gradients to flow efficiently through the network, mitigating the vanishing gradient problem commonly observed in deep models.

The network is composed of an initial convolutional stem followed by four stages of **bottleneck residual blocks**, with filter depths increasing from 256 to 2048 (conv2_x to conv5_x). Each residual block contains a series of 1×1, 3×3, and 1×1 convolutions, enabling effective dimensionality reduction and hierarchical feature learning. ResNet-50 models are often **pretrained on large-scale datasets such as ImageNet**, providing robust initial feature representations that can be fine-tuned for specific tasks like land-cover classification. When integrated as the encoder in a U-Net configuration (forming a **ResU-Net**), it combines the spatial precision of U-Net's decoder with the deep semantic understanding of ResNet's encoder, enhancing segmentation performance and generalization across diverse landscapes (He et al., 2016).

2.2.4 Hyperparameter optimization

Prior to training, the dataset was divided into training and testing subsets using stratified sampling to preserve the proportional representation of all land-cover categories. Stratification was based on the dominant class within each image tile, ensuring balanced class distribution across both subsets. The training data were used to optimize model weights, while the validation data provided an unbiased measure of model generalization. This process was initially applied to the Mahottari dataset, which served as the base dataset for model training. The network was compiled using the Adam optimizer (Kingma & Ba, 2014) with a base learning rate of $1*10^{-4}$, and a decay constant of 10^{-4} was introduced to gradually reduce the learning rate as training progressed, promoting stable convergence. To further refine this process, a ReduceLROnPlateau callback was employed, automatically reducing the learning rate when validation loss plateaued, thereby facilitating more efficient optimization.

Hyperparameter optimization played a crucial role in achieving an optimal trade-off between computational efficiency, segmentation accuracy, and training stability. Three tile sizes (128*128, 256*256, and 512*512) pixels were tested to evaluate their influence on spatial context and computational demand. Similarly, various batch sizes (4, 8, and 16) were experimented with to determine optimal memory–convergence trade-offs. The dropout rate was fine-tuned between 0.05 and 0.1, where 0.1 provided slightly better regularization. The learning rate was dynamically adjusted using the ReduceLROnPlateau scheduler, which reduced the learning rate by half whenever the validation loss stagnated, ensuring smoother convergence. Furthermore, early stopping was used to terminate training once no further improvement in validation performance was observed, preventing overfitting and unnecessary computation (Bergstra & Bengio, 2012; Akiba et al., 2019).

2.2.5 Cross-area fine-tuning

Cross-area fine-tuning (CAFT) is a transfer learning approach that adapts a deep learning model trained in one geographic region (Mahottari) to another similar region (Dhanusha) with limited labeled data. The model initially learns general spatial and spectral features such as edges, shapes, vegetation textures, water bodies, and built-up patterns, which are largely transferable across regions. When applied to a new dataset, differences in spectral response, texture, and illumination may arise, but the pre-trained model provides a strong starting point. This leads to faster convergence, reduced risk of overfitting, and improved classification performance, especially when the target dataset is small or less diverse (Pan & Yang, 2010; Shin et al., 2016).

The fine-tuning process is carried out in two phases to effectively adapt the model to the Dhanusha dataset. In the first phase, a significant portion of the network (mainly the encoder) is frozen, and only the decoder is trained, allowing the model to map previously learned generic features to new land cover classes without losing important spatial representations. In the second phase, all layers are unfrozen and trained together using a smaller learning rate, enabling subtle adjustments across the entire network. This controlled update helps the model adapt to local variations such as soil color, vegetation reflectance, and urban patterns, while preserving the valuable knowledge learned from the source region.

2.2.6 Performance evaluation

The evaluation focuses on several quantitative metrics -Accuracy, Precision, Recall, F1-Score, as well as qualitative assessments through visual inspection of output maps and confusion matrices. Together, these approaches help determine whether the model is robust enough for operational LULC monitoring and decision-making applications in the Dhanusha district.

Accuracy measures the overall proportion of correctly classified pixels relative to the total number of pixels in the validation dataset. It provides a general sense of model performance but can be misleading in cases where certain land cover classes dominate the dataset (Powers, 2011). To address this, Precision and Recall are calculated to evaluate class-specific performance. Precision quantifies the percentage of correctly predicted pixels among all pixels predicted for a particular class, reflecting the model's ability to avoid false positives (Sokolova & Lapalme, 2009). Recall, on the other hand, measures the percentage of correctly identified true class pixels out of all actual pixels of that class, indicating how effectively the model detects specific features such as forest, water, or urban areas (Sokolova & Lapalme, 2009). Together, they provide insight into both the reliability and sensitivity of the model's predictions.

The F1-Score, computed as the harmonic mean of Precision and Recall, balances the trade-off between the two metrics, making it particularly useful for imbalanced datasets where some land cover types are underrepresented (Sokolova & Lapalme, 2009).

2.2.7 LULC map preparation

Finally, the optimized models generate comprehensive **LULC classified tiles** providing actionable insights for agricultural planning, flood risk management, and urban development across both districts. The workflow's geographic specialization - training on Mahottari's landscapes and fine-tuning for Dhanusha's unique conditions ensures high accuracy while demonstrating how residual connections in ResUNet improve feature learning for complex terrain compared to standard U-Net implementations.

3. RESULTS AND DISCUSSIONS

3.1 U-net segmentation

This study implemented a U-Net–based semantic segmentation framework for Land Use Land Cover (LULC) classification using a cross-area fine-tuning approach, where a model trained on Mahottari district was adapted and evaluated on Dhanusha district to assess spatial transferability and generalization capability (Tuia et al., 2016). Five LULC classes were classified using both RGB Sentinel imagery and composite spectral indices, which improved feature representation by capturing vegetation, moisture, soil, and built-up characteristics (Stoian et al., 2019). The overall accuracy for each RGB and composite is presented in *Table 3*.

For RGB-based segmentation, the U-Net model was trained using multiple tile sizes and batch sizes with the Adam optimizer, dropout regularization, and L2 weight decay to reduce overfitting (Kingma & Ba, 2014; Srivastava et al., 2014). Results showed that larger tile sizes provided better spatial context for segmentation, with the 512×512 tile size achieving the highest accuracies of 0.95 in Mahottari and 0.94 after fine-tuning on Dhanusha. Cropland and Tree Cover classes achieved the highest precision and F1-scores, while Water Bodies and Built-Up classes showed lower recall due to spectral mixing and smaller object sizes. *Figure 3* and *Figure 4* shows training and validation curves indicating stable convergence and effective overfitting control (Ronneberger et al., 2015). *Table 4* Per class precision,recall and F1 score for each class for RGB and composite.

For segmentation using composite spectral indices, the same U-Net architecture and cross-area fine-tuning strategy were applied. The integration of spectral indices improved the discrimination of complex land-cover classes by combining spectral and spatial information (Stoian et al., 2019). Adaptive learning-rate scheduling using ReduceLROnPlateau enhanced training stability. The best performance was achieved with the 256×256 tile size, reaching accuracies of 0.95 in Mahottari and 0.93 in Dhanusha. *Table 5* presents U-Net segmentation prediction for different tiles of Dhanusha district.

3.2 ResNet-50 segmentation

This study employed a ResNet-50–based semantic segmentation framework for Land Use Land Cover (LULC) classification using a cross-area fine-tuning approach. The model was trained on Mahottari district data and fine-tuned on Dhanusha district to evaluate spatial transferability and generalization. Five LULC classes -Water, Cropland, Built-Up, Tree Cover, and Others were classified using both RGB imagery and composite spectral indices. Training utilized different tile sizes (128×128, 256×256, and 512×512) and batch sizes, while optimization included ReduceLROnPlateau learning-rate scheduling, dropout regularization, L2 regularization, and early stopping to minimize overfitting.

For RGB-based segmentation, larger tile sizes improved performance by providing greater spatial context for identifying complex land-cover patterns. The best results were achieved with 512×512 tiles, producing accuracies of 0.96 in Mahottari and 0.95 after fine-tuning on Dhanusha that is presented in *Table 6*. Although slight performance reduction occurred due to regional spectral differences, the high accuracy demonstrated the effectiveness of cross-area fine-tuning for improving model generalization in regions with limited labeled data (Tuia et al., 2016). Cropland, Tree Cover, and Others classes achieved high precision, recall, and F1-scores above 0.95, while Water and Built-Up classes showed lower recall because of spectral confusion with surrounding land-cover types as shown in *Table 7*. Training and validation curves shows stable convergence and controlled overfitting as presented in *Figure 5* and *Figure 6*.

The study also integrated composite spectral indices into the ResNet-50 encoder within a U-Net segmentation framework. Residual connections enhanced deep spatial-spectral feature extraction and reduced gradient degradation during training. The 256×256 tile size achieved the best performance, with accuracies of 0.95 in Mahottari and 0.94 in Dhanusha, suggesting that moderate tile sizes provide a better balance between spatial context and training stability (Hoeser & Kuenzer, 2020). The integration of spectral indices significantly improved segmentation robustness and transferability across geographically distinct regions for remote sensing–based LULC classification (Persello et al., 2022). *Table 8* presents ResNet-50 segmentation prediction for tiles of Dhanusha district.

3.3 Discussions

This study examined the comparative performance of deep learning architectures, U-Net and ResNet-50 for semantic segmentation of LULC from Sentinel-2 imagery, emphasizing cross-area fine-tuning, hyperparameter optimization, and spectral index fusion. The findings reveal that both architectures exhibit strong potential for automated LULC mapping across agro-ecologically similar districts in Nepal, with ResNet-50 outperforming U-Net in overall accuracy and generalization. However, the performance gap narrows when composite spectral indices are introduced, highlighting the importance of input spectral richness in deep feature extraction.

The analysis across tile sizes (128×128, 256×256, 512×512) revealed a consistent relationship between spatial context and model performance. Both U-Net and ResNet-50 achieved their best balance of accuracy and class sensitivity at 256×256, with larger tiles (512×512) slightly improving overall accuracy but degrading recall for minority classes such as Water and Built-Up. This supports findings by (Kattenborn et al., 2021) and (Mboga et al., 2017), who emphasized that intermediate receptive fields preserve both fine spatial detail and global context. Excessively small tiles limit contextual awareness, while excessively large one's blur class boundaries, particularly in heterogeneous landscapes.

Hyperparameter optimization was also critical for convergence stability. Learning rate 10^{-4} using **ReduceLROnPlateau scheduling, dropout rate of 0.1**, **L2 regularization of 10^{-4}** produced the most robust results. These configurations align with the optimization strategies described by (Goyal et al., 2017), where moderate dropout and controlled learning rates reduce overfitting in convolutional networks(Prechelt, 2012).

A major finding of this study was the significant improvement achieved when spectral indices were incorporated as composite inputs. The addition of these indices enhanced the discriminative capability of both architectures, particularly in separating Built-Up and Water classes from spectrally similar surfaces such as bare soil or vegetation. This improvement mirrors conclusions from (Stoian et al., 2019) and (Zhang et al., 2019), who reported that spectral index fusion provides complementary biophysical information such as vegetation vigor, moisture content, and urban reflectance thereby reducing class confusion.

For instance, NDWI and NDBI improved separability between urban and aquatic features (Xu, 2006: Zha et al., 2003; McFeeters, 1996), while NDVI and SAVI stabilized vegetation classification by mitigating soil background noise (Huete, 1988; Tucker, 1979). Consequently, the composite-based ResNet-50 achieved 0.94 overall accuracy and an average F1-score of 0.86, outperforming its RGB counterpart by approximately +0.02 in F1. This trend validates findings by (Phiri et al., 2020) and (Talukdar et al., 2020), who demonstrated that integrating spectral indices substantially enhances class-level precision in multispectral segmentation.

The U-Net architecture demonstrated strong performance (OA = 0.92–0.94) with robust generalization under cross-area fine-tuning. However, ResNet-50 consistently outperformed U-Net across both RGB and composite configurations, attaining peak accuracy of 0.95 with RGB and 0.94 with spectral indices. This superiority arises from ResNet's residual connections, which facilitate deeper feature propagation without gradient vanishing (He et al., 2016). The skip connections in ResNet-50 allowed for hierarchical learning of both local textures and global spatial dependencies, crucial for distinguishing spectrally overlapping classes such as Built-Up and Cropland. Similar conclusions were reported by (Audebert et al., 2016) and (Maggiori et al., 2017).

Cross-Area Fine-Tuning played a pivotal role in enhancing generalization between the Mahottari (source) and Dhanusha (target) districts. Models pretrained on Mahottari imagery and fine-tuned on Dhanusha exhibited high transferability for vegetation-dominated classes, Cropland and Tree Cover achieving recall values above 0.95. This aligns with the observations of (Yosinski et al., 2014), who highlighted that feature reuse across similar domains accelerates convergence and mitigates overfitting. However, transfer performance declined for Water and Built-Up classes due to spatial heterogeneity and class imbalance, a common challenge in land-cover modeling (Castelluccio et al., 2015).

Misclassification between Water and Agriculture classes was mainly caused by similar spectral reflectance in shallow water, moist soil, and dry agricultural land (Qiu et al., 2020; Wurm et al., 2019). NDWI and EVI helped reduce this confusion by improving spectral contrast. In this study, bare land was mostly labeled as Cropland because it represented fallow farmland, while urban bare surfaces were classified as Built-Up. This created spectral ambiguity between Cropland and Built-Up classes, while shallow and sediment-rich water bodies reduced Water classification accuracy in heterogeneous rural landscapes.

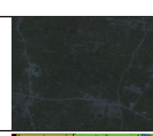
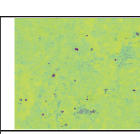
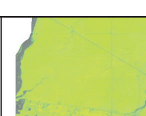
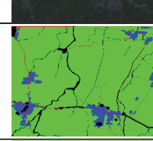
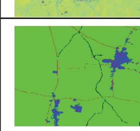
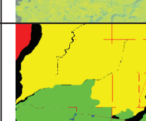
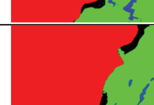
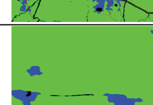
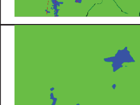
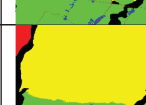
Table 3: Overall accuracy U-net segmentation

Tile Size	Batch Size	Epochs	dropout	Overall Accuracy(RGB)		Overall Accuracy(Composite)	
				Model Trained and Tested on Mahottari	Model Fine Tuned on Dhanusha	Model Trained and Tested on Mahottari	Model Fine Tuned on Dhanusha
128*128	16	60	0.1	0.93	0.91	0.93	0.92
256*256	8	60	0.1	0.93	0.92	0.95	0.93
512*512	4	60	0.1	0.95	0.94	0.92	0.90

Table 4: Accuracy metrics U-net segmentation

S.N	Class	U-Net (RGB)			U-Net (Composite)		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score
1	Water Bodies	0.65	0.57	0.61	0.74	0.64	0.69
2	Cropland	0.92	0.96	0.94	0.93	0.96	0.95
3	Built-Up	0.76	0.42	0.54	0.76	0.52	0.62
4	Tree Cover	0.92	0.97	0.94	0.93	0.98	0.96
5	Others	1.0	0.94	0.97	1.0	0.95	0.97

Table 5: U-net segmentation prediction

Tiles				
Masks				
Prediction				

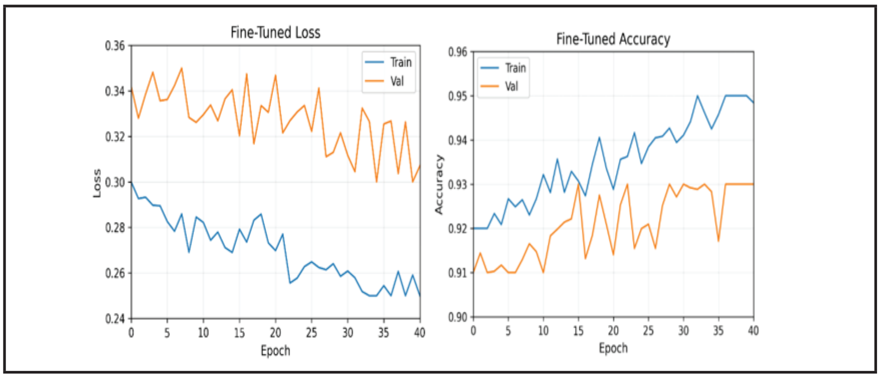


Figure 3 :Fine-Tuned loss and accuracy (RGB) U-net

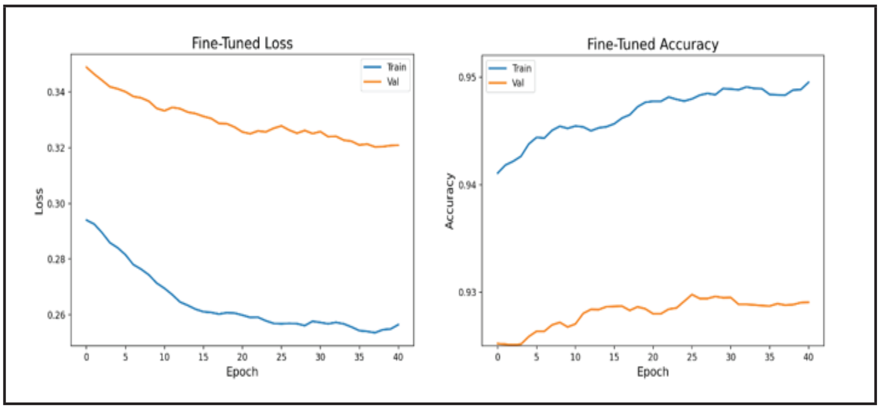


Figure 4 :Fine-tuned loss and accuracy (Composite) U-Net

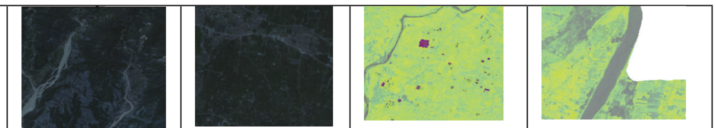
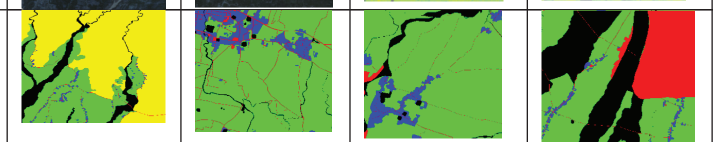
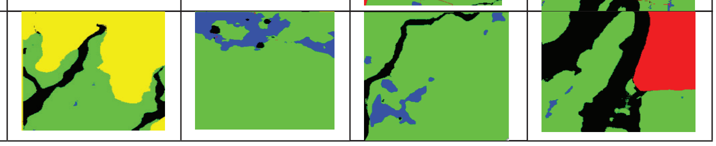
Table 6: Overall accuracy ResNet-50 segmentation

Tile Size	Batch Size	Epochs	dropout	Overall Accuracy(RGB)		Overall Accuracy(Composite)	
				Model Trained and Tested on Mahottari	Model Fine Tuned on Dhanusha	Model Trained and Tested on Mahottari	Model Fine Tuned on Dhanusha
128*128	16	60	0.1	0.93	0.92	0.93	0.92
256*256	8	60	0.1	0.95	0.93	0.95	0.94
512*512	4	60	0.1	0.96	0.95	0.89	0.87

Table 7: Accuracy metrics Res-net-50 segmentation

		ResNet-50 (RGB)			**ResNet-50 (Composite)**		
S.N	**Class**	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
1	Water Bodies	0.85	0.56	0.67	0.82	0.68	0.74
2	Cropland	0.92	0.98	0.95	0.94	0.98	0.96
3	Built-Up	0.79	0.52	0.63	0.77	0.60	0.67
4	Tree Cover	0.95	0.98	0.96	0.95	0.99	0.97
5	Others	1.0	0.98	0.99	1.0	0.95	0.97

Table 8: ResNet-50 segmentation prediction

Tiles	
Masks	
Prediction	

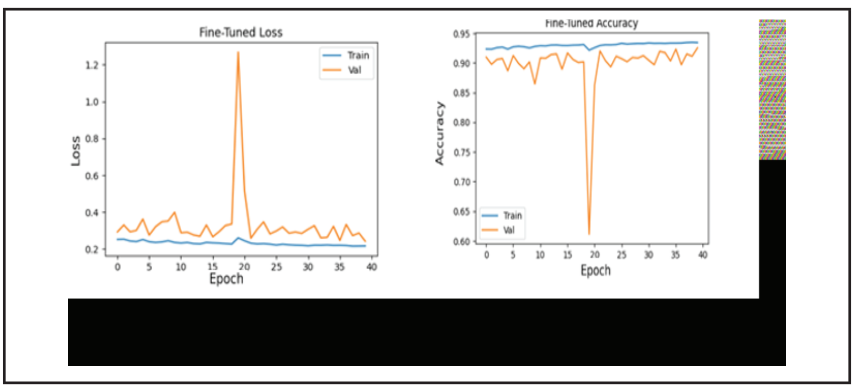


Figure 5:Fine-tuned loss and accuracy (RGB) ResNet-50

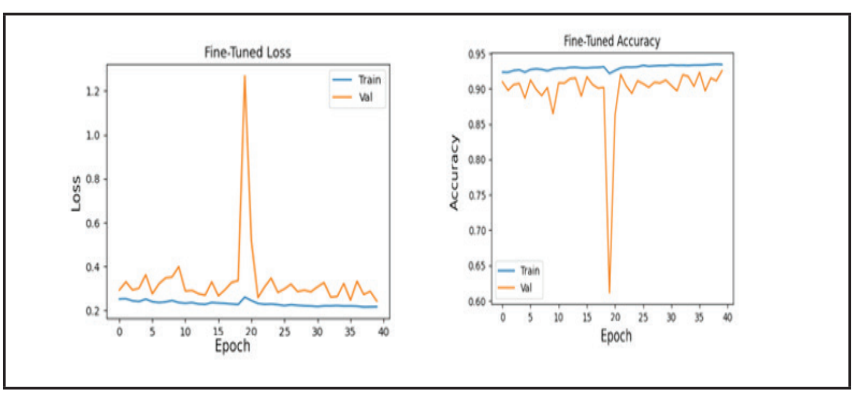


Figure 6: Fine-tuned loss and accuracy (Composite) ResNet-50

4. CONCLUIONS AND RECOMMENDATIONS

This study developed an optimized deep learning framework for Land Use/Land Cover (LULC) classification in the Terai districts of Mahottari and Dhanusha, Nepal, by integrating cross-area fine-tuning, hyperparameter optimization, and spectral-index fusion using Sentinel-2 imagery. Among the tested models, ResNet-50 outperformed U-Net, achieving an overall accuracy of 0.94 and an average F1-score of 0.86, while U-Net achieved an overall accuracy of 0.93 and an F1-score of 0.84. The inclusion of spectral indices such as NDVI, NDWI, and NDBI improved class separability, particularly for vegetation and water-related classes. The optimal training configuration (256×256 tile size, batch size 8, dropout 0.1, and learning rate of 10^{-4}) provided stable convergence and balanced computational efficiency.

The study concludes that combining cross-area fine-tuning with spectral-index integration and optimized training strategies offers a scalable and data-efficient solution for LULC mapping in data-scarce regions. However, limitations such as the use of single-date imagery, limited field data for minority classes, and computational constraints affected model generalization and exploration. Future research should incorporate automated hyperparameter optimization methods, additional data sources such as SWIR bands, topographic information, and multi-temporal imagery, as well as advanced architectures including ResUNet, DeepLabv3+, and Transformer-based models to further improve classification accuracy and large-scale environmental monitoring.

REFERENCES

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). *Optuna: A Next-generation Hyperparameter Optimization Framework* (Version1)

Audebert, N., Saux, B. L., & Lefèvre, S. (2016). *Semantic Segmentation of Earth Observation Data Using Multimodal and Multi-scale Deep Networks* (Version 1).

Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, *13*(10), 281–305.

Castelluccio, M., Poggi, G., Sansone, C., & Verdoliva, L. (2015). *Land Use Classification in Remote Sensing Images by Convolutional Neural Networks* (Version 1).

Diakogiannis, F. I., Waldner, F., Caccetta, P., & Wu, C. (2020). ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *162*, 94–114.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, *120*, 25–36.

Gevaert, C. M., & García-Haro, F. J. (2015). A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. *Remote Sensing of Environment*, *156*, 34–44.

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, *58*(3), 289–298.

Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., Tulloch, A., Jia, Y., & He, K. (2017a). *Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour* (Version 2).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1–2), 195–213.

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*(3), 295–309.

Hoeser, T., & Kuenzer, C. (2020). Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. Remote Sensing, 12(10), 1667.

Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of*

Photogrammetry and Remote Sensing, *173*, 24–49.

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization* (Version 9).

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing, 152, 166–177.*

Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(2), 645–657.

Mboga, N., Persello, C., Bergado, J., & Stein, A. (2017). Detection of Informal Settlements from VHR Images Using Convolutional Neural Networks. *Remote Sensing*, *9*(11), 1106.

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(3), 247–259.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering, 22*(10), 1345–1359.

Paudel, B., Zhang, Y., Li, S., & Liu, L. (2018). *Spatiotemporal changes in agricultural land cover in Nepal over the last 100 years*. *28*.

Persello, C., Wegner, J. D., Hansch, R., Tuia, D., Ghamisi, P., Koeva, M., & Camps-Valls, G. (2022). Deep Learning and Earth Observation to Support the Sustainable Development Goals: Current approaches, open challenges, and future opportunities. *IEEE Geoscience and Remote Sensing Magazine, 10(2), 172–200.*

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies, 2*(1), 37–63.

Prechelt, L. (2012). Early Stopping—But When? In G. Montavon, G. B. Orr, & K.-R. Müller (Eds.), *Neural Networks: Tricks of the Trade* (Vol. 7700, pp. 53–67).

Qiu, C., Schmitt, M., Geiß, C., Chen, T.-H. K., & Zhu, X. X. (2020). A framework for large-scale mapping of human settlement extent from Sentinel-2 images via fully convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, *163*, 152–170.

Richards, J. A., & Jia, X. (1999). *Remote Sensing Digital Image Analysis: An Introduction*. Springer.

Rouse, J., Haas, R. H., Schell, J. A., & Deering, D. (1973). *Monitoring vegetation systems in the great plains with ERTS.*

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging, 35*(5), 1285–1298.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from

Overfitting. *Journal of Machine Learning Research*, *15*(56), 1929–1958.

Stoian, A., Poulain, V., Inglada, J., Poughon, V., & Derksen, D. (2019). Land Cover Maps Production with High Resolution Satellite Image Time Series and Convolutional Neural Networks: Adaptations and Limits for Operational Systems. *Remote Sensing*, *11*(17), 1986.

Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal, S., Liou, Y.-A., & Rahman, A. (2020). Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote Sensing*, *12*(7), 1135

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150.

Tuia, D., Persello, C., & Bruzzone, L. (2016). Domain Adaptation for the Classification of Remote Sensing Data: An Overview of Recent Advances. *IEEE Geoscience and Remote Sensing Magazine*, *4*(2), 41–57.

Uddin, K., Shrestha, H. L., Murthy, M. S. R., Bajracharya, B., Shrestha, B., Gilani, H., Pradhan, S., & Dangol, B. (2015). Development of 2010 national land cover database for the Nepal. *Journal of Environmental Management*, *148*, 82–90.

Wurm, M., Stark, T., Zhu, X. X., Weigand, M., & Taubenböck, H. (2019). Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, *150*, 59–69.

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, *27*(14), 3025–3033.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). *How transferable are features in deep neural networks?*

Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, *24*(3), 583–594.

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., & Atkinson, P. M. (2019). Joint Deep Learning for land cover and land use classification. *Remote Sensing of Environment*, *221*, 173–187.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 8–36.

Author's Information

Name	:	Hiralal Pardhe
Academic Qualification	:	M.E Geoinformatics
Organization	:	Survey Office Kalanki
Current Designation	:	Survey Officer