International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I6P119

Volume 73 Issue 6, 225-237, June 2025


*Original Article*

# Integration of Sentinel-2 Semantic Segmentation Results with Spectral Indices for Forest Change Detection

Shilpa P. Pimpalkar[1], Sai Madhavi D[2]

*[1,2]Rao Bahadur Y. Mahabaleswarappa Engineering College, Ballari, Visvesvaraya Technological University, Belagavi, Karnataka, India.*
*[1]AISSMS Institute of Information Technology, Pune, Maharashtra, India.*

*[1]Corresponding Author : shilpa.pimpalkar@gmail.com*

Received: 28 February 2025 Revised: 08 May 2025 Accepted: 29 May 2025 Published: 28 June 2025

***Abstract -*** *Deforestation presents significant challenges to biodiversity, high conservation-value species, global warming, and climate change. Recent developments in satellite image processing using deep learning and computer vision techniques have greatly improved methods for monitoring forest cover changes. This paper introduces a novel hybrid approach named AResU-Net, an attention-based residual-U-Net for forest change detection. The model generates a binary mask for the given input image leveraging Red (R), Green (G), and Blue (B) bands, along with the Near-Infrared (NIR) band from Sentinel-2 imagery. Spectral indices NDSI and NDVI identify snow/ice masks from the input image. Snow masks are removed from predicted masks to eliminate the possibility of incorrect deforestation detection due to seasonal variations, especially in winter. The effectiveness of the proposed model is validated by experimental results that demonstrate substantial improvements in key performance metrics: "Accuracy" - 0.964782, "Precision" - 0.946866, "Recall" - 0.968095, "F1 Score" - 0.957363, and "Mean Intersection over Union (mIOU)" - 0.929982. The evaluation metrics, complemented by visual analyses, indicate a strong correlation, confirming the model's effectiveness in accurately detecting forest changes. The performance assessment was conducted using a diverse array of validation images, including randomly selected .tif images of varying sizes sourced from Google Earth Engine in regions of Nepal, with an integration of the snow index. The AResU-Net model represents a significant advancement in automated image segmentation methodologies, contributing to environmental conservation efforts by effectively monitoring deforestation trends over time.*

***Keywords -*** *Afforestation, Deforestation, Forest change, Residual, Sentinel.*

## 1. Introduction

Deforestation pervasively influences ecosystems and biodiversity worldwide, emphasizing the importance of comprehensive monitoring and mitigation activities. Aside from natural calamities, increased human activities such as agriculture, mining, and logging are the chief causes of deforestation. An effective monitoring strategy is required to prompt intervention against unlawful human activities and analyze forest cover changes after disasters. Recent research in deep learning and computer vision has underlined the importance of image segmentation in forest change detection. Forest change identification can be automated using high-resolution satellite data and semantic segmentation techniques. Deep learning combined with remote sensing creates a potent technique for accurately segmenting and detecting deforestation. Compared to traditional methods, deep learning-based image segmentation offers significant advantages, including improved accuracy and efficiency in identifying deforested areas. Additionally, it has the potential to fully automate the detection process, making mapping faster and more streamlined. However, this approach also presents challenges, such as the need for extensive training data and the complexity of analyzing diverse and heterogeneous forest landscapes, which are crucial for evaluating the performance of the deep learning model. Despite the recent deep learning-based implementations in this area, challenges still need to be addressed- i) forest change detection (afforestation/deforestation) using heterogeneous forest landscapes. ii) evaluation of the percentage of afforestation and deforestation ii) addressing the issues of seasonal variations due to snow cover, which may lead to incorrect deforestation detection. This paper addresses these challenges by presenting AResU-*Net,* a novel semantic segmentation that uses attention gates and residual blocks with U-Net for semantic segmentation.

## 2. Related Work

Given the latest discoveries in the field, a thorough investigation has been carried out. A review of the literature outlining recent developments in this field is included in this study, along with dataset analysis and an examination of different forest regions. Using the stochastic gradient descent (SGD) optimizer, K. Isaienkov et al. demonstrated experimental results using the U-Net model combined with ResNet-50 and U-Net 101. The authors created the Sentinel-2 dataset for their research, which sought to identify deforestation in specific regions of Ukraine. They underlined the need for more datasets to be developed for training and future improvements in dataset accuracy [1]. A potential hybrid semantic segmentation model that combines the attention U-net model with the residual connection was proposed by K. Kalinaki et al. [2]. Despite requiring more processing time than the Attention U-Net and the standard U-Net, the model's greater accuracy revealed its potential [2]. F. Pelletier et al. [3] created an algorithm trained on tree masks to distinguish tree crowns from other land cover types.

The algorithm utilized Shortwave Infrared (SWIR) and Normalized Burn Ratio (NBR) for change detection. Pacheco-Pascagaza et al. [4] emphasized using the Sentinel-2 dataset (10m resolution), which can be effectively used for deforestation detection to enable the logging of individual large tropical trees. Jamali et al. proposed a hybrid semantic segmentation model that integrated components of HetConv, U-Net, attention gates, and Vision Transformers (ViT). The proposed solution has shown significant improvement in forest change detection compared with existing findings [5]. David John et al. conducted a comparative analysis of the results obtained using Attention U-Net against “U-Net”, “Residual U-Net”, “ResNet50-SegNet”, and “FCN32-VGG16” across three distinct datasets: “three-band Amazon, four-band Amazon, and Atlantic Forest”. Their findings indicated that the Residual Attention U-Net, while containing a greater number of parameters than the Attention U-Net, may result in an extended training duration but holds the potential for improving upon the performance of the Residual U-Net[6]. D. Lee and Y. Choi et al. [7] compared several U-Net family networks, including “U-Net, R2U-Net, Attention U-Net, Attention R2U-Net, and Nested U-Net”. The assessment focused on two key evaluation metrics: “F1-score” and “Intersection over Union (IoU)”.

The findings revealed that the Attention U-Net demonstrated superior predictive performance for both metrics, utilizing satellite imagery from “Sentinel-1”, “Sentinel-2”, and “Landsat 8”. J. B. Kilbride et al. [8] employed Google Earth Engine and Google AI platforms to detect forest cover changes. They underscored the necessity of refining their training datasets within these platforms to enhance detection accuracy. Furthermore, the authors noted that classifiers must exhibit robustness against noise resulting from seasonal variations in the data, with Landsat imagery serving as the analytical basis. M. Kaselimi et al. [9] introduced the Vision Transformer Model for change detection with data source- “https://www.kaggle.com/c/planet-understanding-the-amazon-from-space/data”. S. V. Lim et al. [10] explored attention-based semantic segmentation utilizing Landsat 8 imagery for change detection. Future directions are suggested to i)contribute to developing a dataset including seasonal variations and ii) improve the architecture of attention-based semantic segmentation models and evaluate their effectiveness. Studies advised using high-performance GPUs or investigating alternative pipelining strategies to facilitate processing extensive datasets within constrained GPU memory. Another approach proposed was to utilize datasets with slightly reduced resolutions to alleviate memory constraints. P. Tovar et al. [11] investigated the “Spatial Attention Mechanism (SAM)” and “Channel Attention Mechanism (CAM)” utilizing optical images obtained from the Landsat 8-OLI sensor, characterized by a resolution of 30 meters.

Their experiments sought to assess the relevance and sensitivity of hyperparameter tuning for the loss function and the impact of dual-attention mechanisms (spatial and channel) on predicting deforestation. K. Karaman et al. [12] introduced an attention-based neural network architecture called “SENTINEL-1 BraDD-S1TS”, encompassing approximately 25,000 image sequences depicting deforested and unchanged land within the Brazilian Amazon. I. Md Jelas et al. [13] presented a detailed literature survey mentioning the effectiveness of recent image segmentation techniques ‘U-Net’, ‘DeepLabV3’, ‘ResNet’, ‘SegNet’, and ‘Fully Convolutional Networks (FCN)’. They suggested contributing to developing a ‘ResU-Net’ to utilize skip connections from U-Net with the residual blocks from ResNet.

The suggested integration has the potential to extract finer patterns in deforested areas, with residual blocks contributing significantly to enhanced model precision.D. L. Torres [14] assessed “U-Net, ResU-Net, SegNet, FC-DenseNet” and two variants of “DeepLabv3+” for monitoring Brazilian Amazon deforestation. The study was presented by utilizing Landsat-8 and Sentinel-2 images. The findings indicated that ResU-Net consistently exhibited the most favourable trade-off between accuracy and training and inference times. In contrast, MobileNetV2 and SegNet demonstrated the least effective performance among the evaluated frameworks. This study constitutes a noteworthy advance toward achieving more efficient, objective, and accurate monitoring of the Amazon rainforest. P. J. Soto [15] examined various regions within the Amazon rainforest and the Brazilian Cerrado, utilizing Landsat 8-OLI. The authors devised an unsupervised pseudo-labelling scheme based on the Change Vector Analysis technique to mitigate the challenges posed by the scarcity of ground-truth labelled training samples. T. Andrei [16] presents two principal contributions to the field: 1. The automation of image labelling using a Gaussian Mixture Model (GMM), 2. The development of a U-Net model that minimizes resource requirements, computational time, and model complexity while preserving accuracy.

The study utilizes a Geospatial Defense Intelligence Agency (AIGA) dataset. B. M. Matosak [17] has developed a model integrating Long Short-Term Memory (LSTM) and U-Net. The study utilized Landsat-8 and Sentinel-2 images for detecting deforestation in the Cerrado region. Z. Wang [18] introduces RepDDNet, which is capable of effectively extracting contextual semantic features with Siamese backbone and encoder-decoder architecture. Compared to deep learning models incorporating re-parameterization or segmentation, the proposed RepDDNet demonstrated exceptional performance. Haseeb Azizi [19] employed Random Forests (RF) to estimate Fractional Snow Cover (FSC). The analysis revealed that the RF models performed variably across different elevation zones.This literature survey is centered on research related to forest change detection. The datasets utilized, and the methodologies applied in prior findings are detailed in the accompanying figures: Figure 1 and Figure 2.

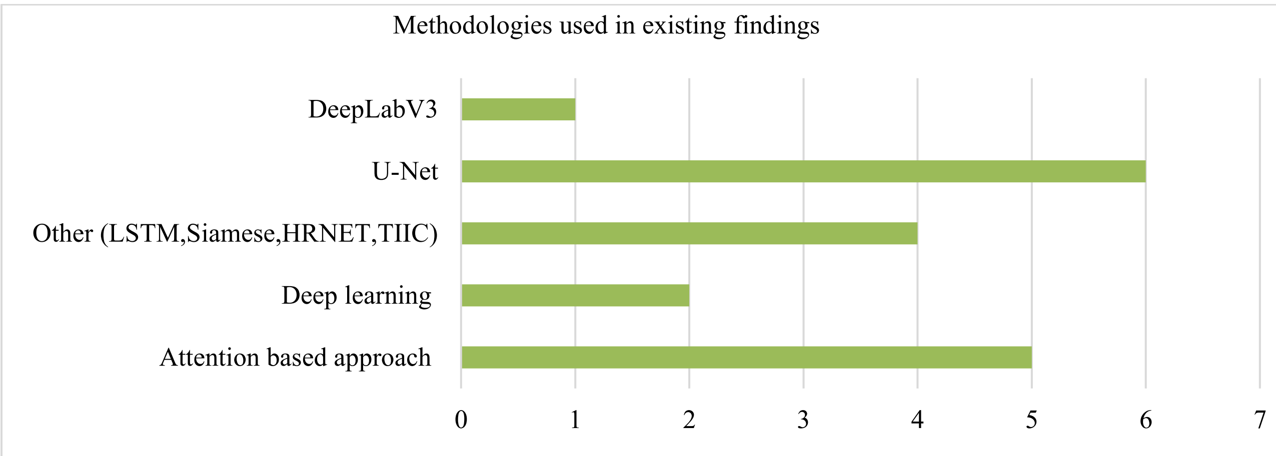



**Fig. 1 Methodologies used in existing findings**

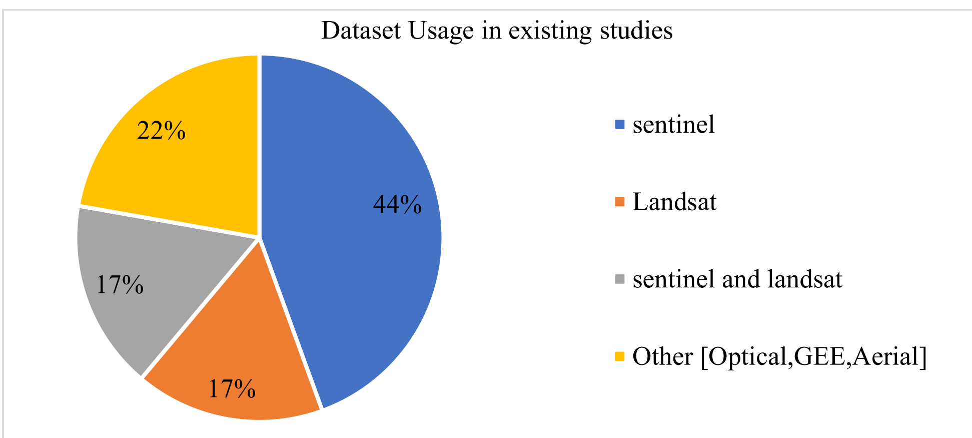



**Fig. 2 Percentage of dataset usage in existing studies**

As shown in Figure 1. U-net is used as a baseline model in most existing findings, and few findings have presented enhanced performance using an attention-based approach. There are very few papers based on DeepLabV3. As shown in Figure 2, 44% of studies used the Sentinel dataset,17% used the Landsat dataset,17% of studies presented results using both Sentinel and Landsat datasets, and 22% of studies presented results using other datasets, e.g. Optical images, images acquired using Google Earth Engine and aerial images.

Key findings based on the literature survey are as follows:

1. Sentinel2 dataset is used in most of the existing findings, and U-Net is used as a baseline model for forest change detection.
2. All the recent findings suggested using attention-based semantic segmentation approaches for change detection. Attention U-Net demonstrates the highest performance among the models evaluated [6, 5, 2, 10, 7, 11].
3. Few studies highlighted future research using the Residual Attention U-Net, which includes a greater number of parameters than the Attention U-Net, as it has the potential to enhance the performance of the Residual U-Net [6, 13].
4. We must address the challenges in processing images with seasonal variations [8].

5. Further studies are needed to utilise diverse datasets to make Amazon rainforest monitoring more efficient [14].
6. A fixed NDSI threshold is needed to identify seasonal variation in forest cover [19].

The following research gaps are identified through the literature survey-

1. Findings that combine the Sentinel-2 dataset with spectral indices for forest cover change detection are unavailable.
2. Existing findings highlighted the need for forest change detection using a semantic segmentation model that combines attention gates and residual blocks to enhance performance.
3. Existing findings highlighted the need for research that combines spectral indices (NDSI, NDVI )with semantic segmentation results to detect seasonal variations in forest cover (e.g. Snow cover).

This study proposes an AResU-Net attention-based residual U-Net semantic segmentation model that combines results with spectral features for forest change detection. The objective of the proposed model is to estimate the percentage of forest change (afforestation/deforestation) and identify snow cover to eliminate the possibility of incorrect detection of deforestation during winter. Details about the material and methods are given in section 3. Section 4 presents the result, and section 5 concludes and presents future directions.

## 3. Materials and Methods

The *AResU-Net* semantic segmentation model is presented as an advanced approach for analysing deforestation and afforestation dynamics. The model is designed to process a pair of input images (.tiff ) sequentially and generate binary temporal masks (M1 and M2). The model is trained using a pair of .tiff files and associated binary mask to produce an output binary mask. In the predicted binary temporal mask, deforested areas are represented in black, while forested areas are depicted in white. The temporal binary masks generated through this process (M1 and M2) are subsequently compared pixel by pixel to facilitate change detection, and change is stored in the resulting mask (RM).

The snow mask is derived from the input images (.tiff) using the “Normalized Difference Vegetation Index (NDVI)” and the “Normalized Difference Snow Index (NDSI)”, which assists in the identification of snow-covered regions. To enhance the accuracy of forest change estimations, snow masks are excluded from the change detection mask (RM), thereby mitigating the risk of erroneously categorizing snow as deforested land.

This research work proposes a system with the ability:-

1) To generate the mask for the unseen input images. The system is proposed to generate the output mask (Geotiff image). Output mask serves as Ground truth /labelled data. In the generated output mask, each pixel is classified as belonging to a forest or not (background).
2) To detect forest changes, i.e., deforestation/afforestation, by comparing two temporal masks.
3) To identify and eliminate snow cover from the change detection mask.

This section further details 3.1 Dataset, 3.2 Proposed Architecture and 3.3. Model Implementation.

### *3.1. Dataset Details*

The dataset contains high-resolution multispectral imagery from the Sentinel-2 Level 2A Satellite. The dataset covers two significant tropical forest regions: the “Amazon Rainforest” and the “Atlantic Forest” in Brazil. The images in the dataset are meant to train a “fully convolutional neural network” designed for semantic segmentation tasks, particularly in monitoring forest cover changes. Each image typically includes multiple spectral bands, such as RGB and Near-Infrared (NIR). The datasets also provide pixel-level annotations to classify areas as forested or deforested, enabling detailed analysis of forest cover changes over time using various deep learning models. The dataset consisted of .tif images of sizes (4, 512, 512). These images were converted to patches of size (4, 128, 128) for model input and were split into ‘training’, ‘validation’, and ‘test datasets’. The models were trained on the training dataset (15155 patches), with validation on the validation dataset (3789 patches).

#### *3.1.1. Dataset Link*

https://zenodo.org/records/4498086#:~:text=This%20database%20contains%20images%20from,%2C%203%2C%202%20and%208

### *3.2. System Architecture*

The layout of the components detailing the system's flow and operations of the proposed forest change detection system is shown in Figure 3. The proposed system is trained using an input image-mask pair to generate the binary mask of the input images indicating the ‘forest’ and ‘non-forest’ regions in the input image. The proposed system demonstrates semantic segmentation of the input image, generating the result as an output mask with a pixel-wise classification indicating the presence or absence of a forest.

The AResU-Net model combines residual connections, attention gates, and an encoder-decoder (U-Net) structure for binary segmentation. Residual blocks extract semantic information from input images, followed by down-sampling with max-pooling layers. In the Encoder structure, five levels of residual blocks progressively extract features, with filter sizes doubling at each level. The decoder structure uses up-sampling layers to restore spatial dimensions and ‘attention gates’ to focus on relevant regions in skip connections. The output layer uses the conv2D layer with sigmoid activation for binary classification, producing a single-channel prediction. The deepest layer, which bridges the encoder and decoder, also uses residual connections.

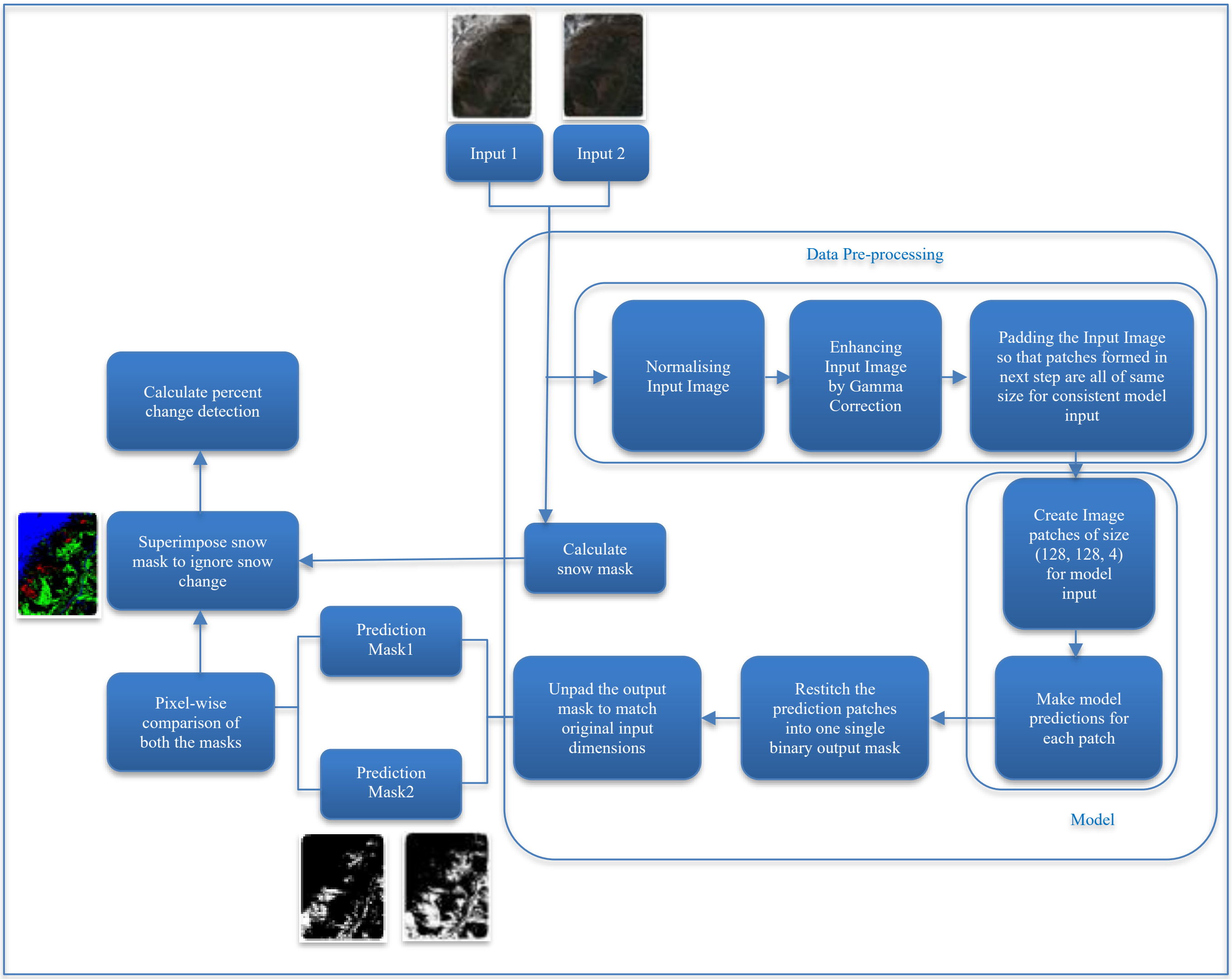



**Fig. 3 System architecture**

### *3.3. Model Implementation*

This section covers details about data pre-processing, major model components, model workflow, change detection and snow identification.

#### *3.3.1. Data Pre-Processing*

This section presents pre-processing of the multispectral image with 4-channel (RGB + NIR) where each pixel is represented by four spectral values 'blue (band 2)', 'green (band 3)', 'red (band 4)', and 'near-infrared (band 8)'. Each pixel will have four values corresponding to the reflectance of bands 2, 3, 4, and 8.

The main preprocessing steps are as follows- Clipping image and image enhancement.

*Clipping Image*

The function is implemented to convert clipped images to NP array as follows:

1) Input Image: GeoTIFF image
2) Clip images to 128 x 128 resolution patches: The input image is clipped into smaller patches of 128 x 128 pixels.
3) Convert input image to numpy arrays.

*Image Enhancement*

An enhanced image is generated by applying a non-linear transformation to adjust contrast using gamma correction. The gamma value is set to 0.5, which can be adjusted for better contrast.

#### *3.3.2. Major Model Components*

The Residual Block is the first important element. Two Conv2D layers with ReLU activation and residual skip connections make up this block. Residual blocks perform well in deep networks because they skip over layers and preserve key properties. The second major component is Attention gates. This component is applied to skip connections to selectively highlight relevant features from encoder layers and ignore less prominent features, which helps in more accurate segmentation. A description of major model components is outlined in Table1.

**Table 1. Major model components**

| Component | Description |
|---|---|
| Residual Blocks | Two Conv2D layers with ReLU activation and skip connections to retain spatial features. |
| Encoder Structure | Five levels of residual blocks with increasing filter sizes, using max-pooling for down-sampling. |
| Bridge Layer | The deepest layer connects the encoder and decoder, utilizing residual connections. |
| Decoder Structure | Up-sampling layers to restore spatial dimensions and integrate skip connections. |
| Attention Gates | Applied to skip connections to highlight relevant features and ignore less prominent features |
| Output Layer | A Conv2D layer with sigmoid activation generates a single-channel binary segmentation mask. |

*Hyperparameters & Training Configuration*

The input shape is (128, 128, 4), where the model accepts 128x128 patches with four channels. As for the number of filters used in Residual Blocks, for progressively deeper layers, the filters used are 64, 128, 256, 512 and 1024. The model is compiled with an AdamW optimizer with a learning rate 1e-3 and weight decay of 1e-5. Binary Cross-Entropy is the loss function, as it is suitable for binary segmentation. A batch size of 4 was chosen to balance training speed and memory requirements. The maximum epoch limit is set to 100. However, an early stopping mechanism is applied through a callback, which monitors validation loss with a patience of 10 to prevent overfitting. Best weights are restored after training is complete.

*3.3.3. Model Workflow: Workflow of the Proposed Model is Given Below*

1. The input image (in .tiff format) is preprocessed and resized to (128, 128, 4).
2. The image is fed into the encoder with residual blocks, progressively extracting features with increasing filters and downsampling.
3. After reaching the bridge layer, the abstract features are passed to the decoder.
4. The decoder upsamples the feature maps, reducing the number of filters progressively.
5. Attention gates refine the features before passing them to the decoder, focusing on the relevant areas.
6. The output layer processes the output from the decoder to produce a binary mask of size (128, 128, 1).
7. The sigmoid activation outputs the probability of each pixel being part of the foreground.
8. The Binary Cross-Entropy loss is computed, and the model is trained with the AdamW optimizer.

By the end of the training process, the model can predict a binary mask for any given .tiff input, effectively performing semantic segmentation.

*3.3.4. Change Detection (Afforestation / Deforestation)*

This function compares the two temporal masks and identifies each pixel as afforested and deforested.

*Steps 1 for Marking Pixel as Afforested and Deforested*

1. It detects deforestation by identifying pixels that changed from white (1) in past_mask to black (0) in current_mask.
2. The detected deforested pixels are assigned a red colour ([255, 0, 0]) in change_mask.
3. It detects afforestation by identifying pixels that changed from black (0) in past_mask to white (1) in current_mask.
4. The detected forested pixels are assigned a green colour ([0, 255, 0]) in change_mask.

*Steps 2 for Calculating Percentage Change*

1. The function calculate_change_percentage computes the percentage of deforested and afforested areas in a given change mask image.
2. It first calculates the total number of pixels in the image.
3. It counts deforested pixels (red: [255, 0, 0]) and afforested pixels (green: [0, 255, 0]) using NumPy operations.
4. The percentages are computed by dividing the respective pixel counts by the total number of pixels and multiplying by 100.
5. The function returns the computed deforestation and afforestation percentages.

*3.3.5. Snow / Ice Cover Detection and Integration with Semantic Segmentation Results*

Detection of the presence of snow helps us minimise the misclassification of afforestation and deforestation areas. Snow can diversly impact models trained to identify forest cover changes, especially when bi-temporal images are captured in different seasons. The snow index helps differentiate snow cover from actual vegetation so the model can focus on identifying true changes in forest cover without being confused by seasonal snow. Thus, the primary goal of snow detection is to exclude snow-covered areas from contributing to change detection. When a model sees a summer image without snow and a winter image with snow, it might interpret the appearance of snow as deforestation. Using a snow index, a snow mask is created in the algorithm, and these areas are identified and masked to ensure they don't influence the final afforestation or deforestation change detection results. A survey of the recent findings is done to calculate snow masks for a given input image. Studies by Poussin C. et al. [20], Wang Y. et al. [21], Wang G. et al. [22]

and He et al. [23] used NDSI to detect the presence of snow. With NDSI>0, NDSI>0.4 is considered as the presence of snow. Wang, Y. et al. [21] and He et al. [23] also used NDVI for snow detection with NDVI>0 no snow presence.

To create a snow mask from the input image, the algorithm leverages NDSI, NDVI and binary dilation. These spectral indices automatically outline landscape features like snow, ice, vegetation and water in Sentinel imagery, where they create an 'index stack' using the three indices sets. The following section presents the computation of NDSI, NDVI and binary dilation.

*Computation of NDSI*

NDSI helps detect areas that might contain snow, thus excluding these areas during change detection. This calculation can be done using NumPy, which supports array-based operations necessary for raster band manipulation.

NDSI=Green-NIR/Green+NIR

Where,

1) NIR is the reflectance value of the Near Infrared Band (Band 8)
2) Green is the reflectance value of the Green band (Band 3)
3) The proposed algorithm uses a threshold of -0.1, i.e. NDWI > -0.1, which helps identify pixels with a reflectance similar to snow.

*Computation of NDVI*

NDVI measures vegetation density and health based on the response in the near-infrared (NIR) and red bands. NDVI values closer to 1 represent denser vegetation. Avoiding areas with low or negative NDVI values is helpful since snow-covered areas can sometimes yield low NDVI, masked with snow index logic. Like NDWI, NumPy handles NDVI computations efficiently by operating on individual bands of the raster.

NDVI=NIR-Red/NIR+Red[21]

Where,

1) NIR is the reflectance value of the Near Infrared Band (Band 8)
2) Red is the reflectance value of the Red band (Band 4)
3) The proposed algorithm uses a threshold of 0.1, i.e.
4) NDVI < 0.1, which excludes pixels with high vegetation density, as snow-covered areas usually have low NDVI.

*Computation of Binary Dilation*

Another technique used for the creation of the snow mask was binary dilation. It is a function from the SciPy library that expands the boundaries of snow-covered areas detected in the snow mask. This operation helps ensure that even the images' isolated or small snow patches are fully included in the mask, reducing potential edge misclassifications. Thus, the binary dilation operation helps produce a more robust semantic segmentation result. Mathematically, binary dilation at each pixel checks if any neighbouring pixels within the structuring element are True (or 1). If at least one neighbouring pixel in the set filter grid 3x3 is set to True, the central pixel is set to True. Thus, for each pixel, the binary dilation operation can be defined as follows:

dilated_mask_matrix(i, j) = max (mask_matrix(i+m, j + n))

Where, m n represents the offsets within the structuring element.

Snow mask calculation steps are listed in Figure 4.

1) Calculate NDSI using bands b3 and b8

   ndsi = (b3 - b8) / (b3 + b8)

2) Calculate NDVI to help filter vegetation areas from the snow mask using B8 and B4

   ndvi = (b8 - b4) / (b8 + b4)

3) Apply NDSI and NDVI thresholds to create a refined snow mask

   snow_mask = (ndsi > -0.1) & (ndvi < 0.1)

4) Apply morphological dilation to expand the detected snow areas. binary_dilation imported from scipy.ndimage.

   snow_mask_dilated = binary_dilation(snow_mask, iterations=1) import

Binary mask generated with 1 for snow, 0 otherwise

**Fig. 4 Steps for snow cover detection using NDVI, NDSI and binary dilation**

The snow masks created using these techniques are then incorporated into the change detection process by masking out these pixels during the afforestation and deforestation change detection process. Thus, using techniques such as NDVI, NDSI and binary dilation and by masking snow areas, the model's performance is enhanced for seasonal images, ensuring only actual afforestation and deforestation changes are highlighted without the interference of temporary snow coverage. By separating out real forest change signals and eliminating the impact of snow on these forecasts, the masking technique improves accuracy. Only actual vegetation-related changes are noted to reduce misinterpretations caused by seasonal variations. The model can perform more robustly across datasets with varying temporal spans because it is less susceptible to seasonal artefacts.

## 4. Results and Discussion

This section presents the findings of the proposed model - AResU-Net Figure 5 shows the binary mask predicted by the proposed model. The predicted binary mask closely aligns with the ground truth, demonstrating the effectiveness of the proposed model. The visual similarity between them indicates high accuracy in segmentation, reinforcing the model's reliability for the given task.

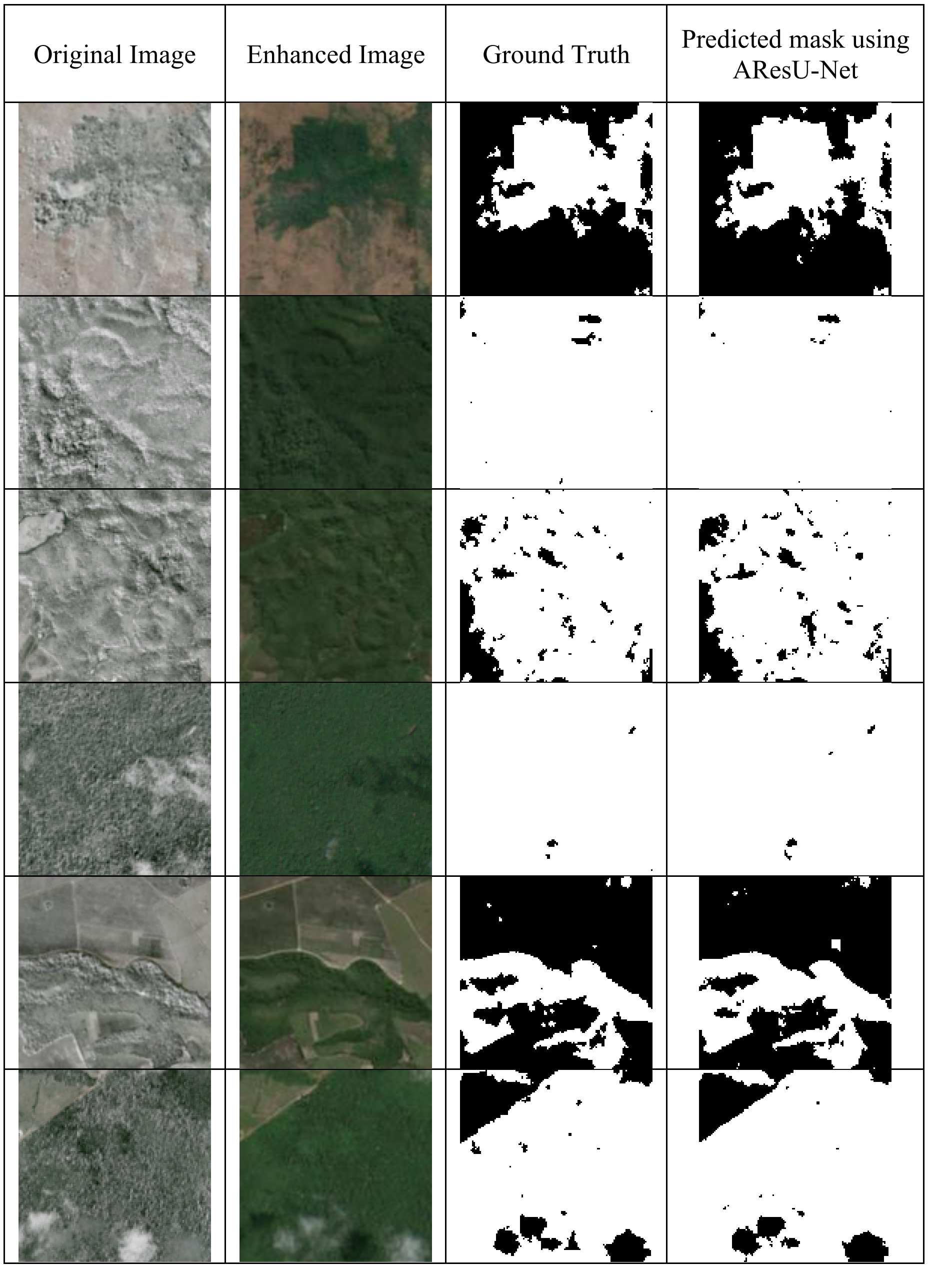



**Fig. 5 Original image, enhanced image, ground truth and predicted mask**

Due to the unavailability of bitemporal images in the test dataset, any two images were taken for the sake of presentation. It was assumed that image1 is the past image, and image2 is the current image.

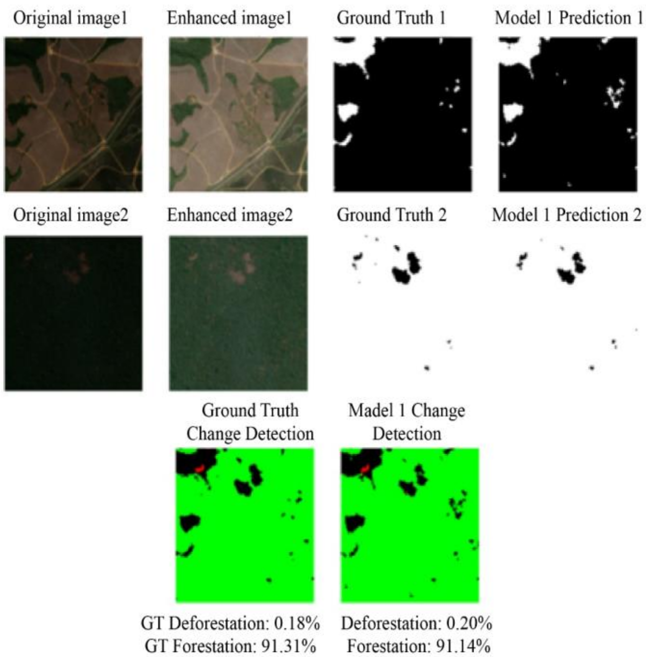
1) The red colour denotes deforestation.
2) The green colour denotes afforestation.
3) Blue colour denotes snow cover.

Further, this section presents the results of the proposed model

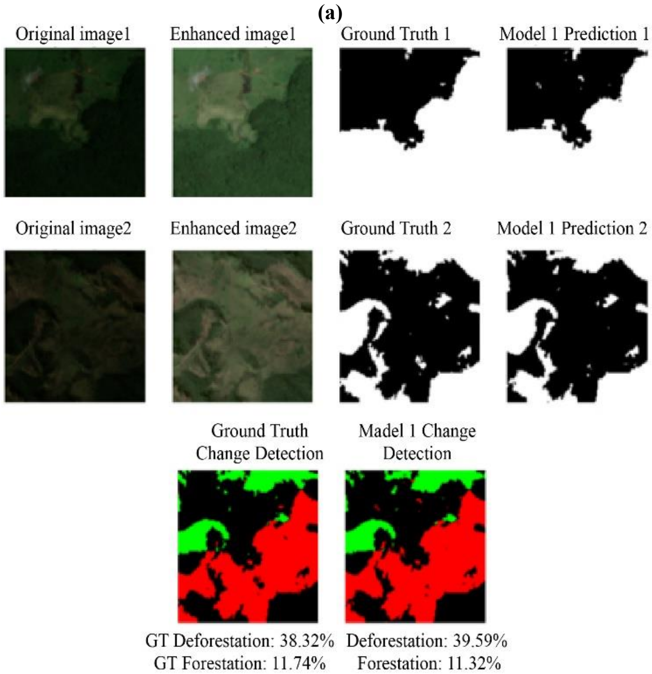
1) Figure 6: Testing using images given in the test dataset
2) Figure 7: Testing using images of large size acquired from Google Earth engine
3) Figure 8: Testing using images from the Nepal region acquired using the Google Earth engine

Figure 6 Shows the result obtained for the images from the test dataset. Each sample includes two input images, enhanced images after preprocessing, their ground truth binary masks, and the binary mask predicted by model 1 (proposed model), and the resultant mask generated is presented along with the ground truth mask.

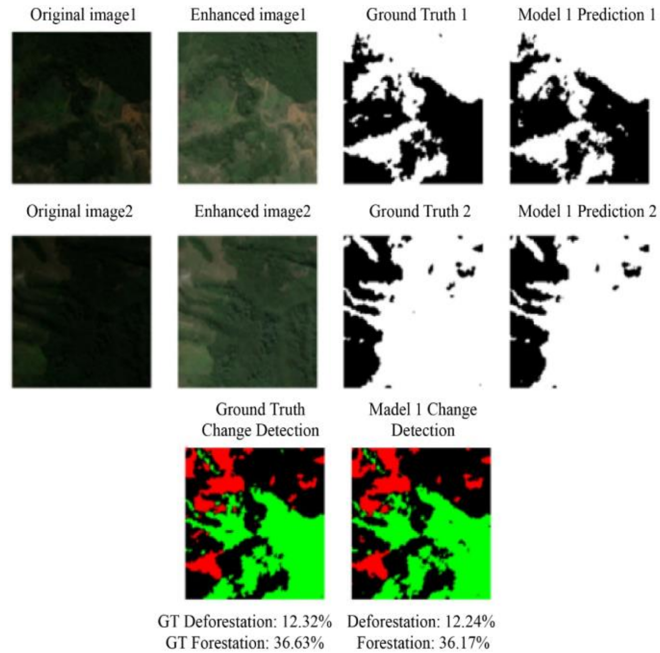
As shown in each sample, ground truth and predicted change detection masks are similar, indicating the model's efficiency in change detection.



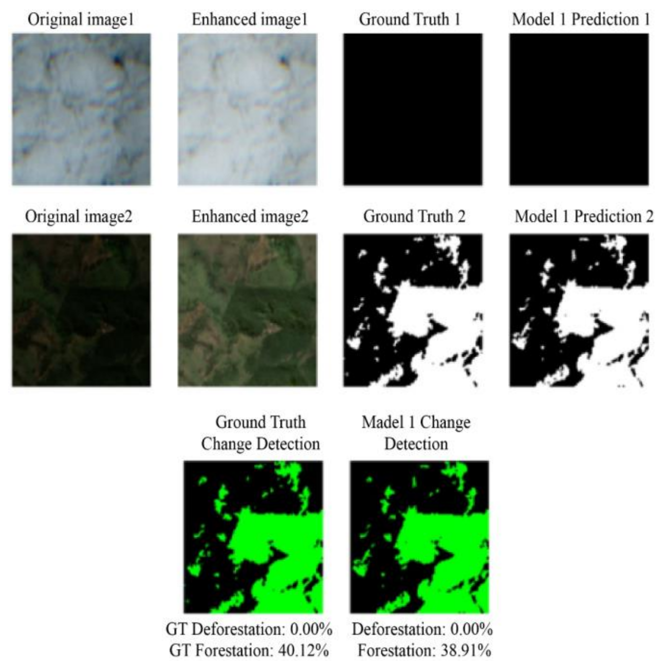

**(a)**




**(b)**




**(c)**

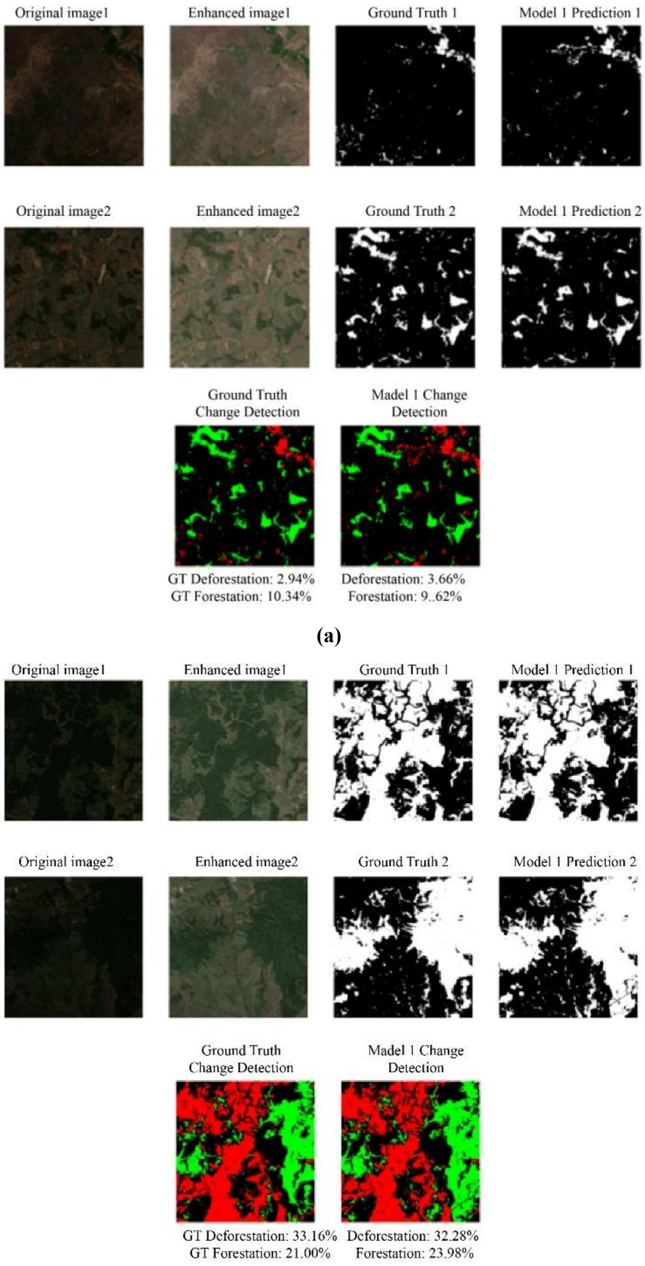



**(d)**

**Fig. 6 Percentage of forest change for test images**

Figure 7 shows the results obtained for large images. The proposed model efficiently processes .tif images of any size, seamlessly handling large-scale inputs by re-stitching the output without losing accuracy.

Its adaptability ensures reliable forest change detection across varying image dimensions, demonstrating robustness in diverse real-world scenarios.




**(b)**

**Fig. 7 Percentage of forest change for images of large size**

Figure 8 shows the results obtained using random .tif images of any size captured from Google Earth Engine from areas of Nepal with integration of snow index.

The proposed model accurately detects changes in afforestation and deforestation while effectively distinguishing snow cover during winter using the integrated snow index. This ensures that seasonal snow is not misclassified as deforestation. Results on random .tif images from Google Earth Engine over Nepal validate the model's high accuracy in change detection, demonstrating its reliability in forest monitoring.

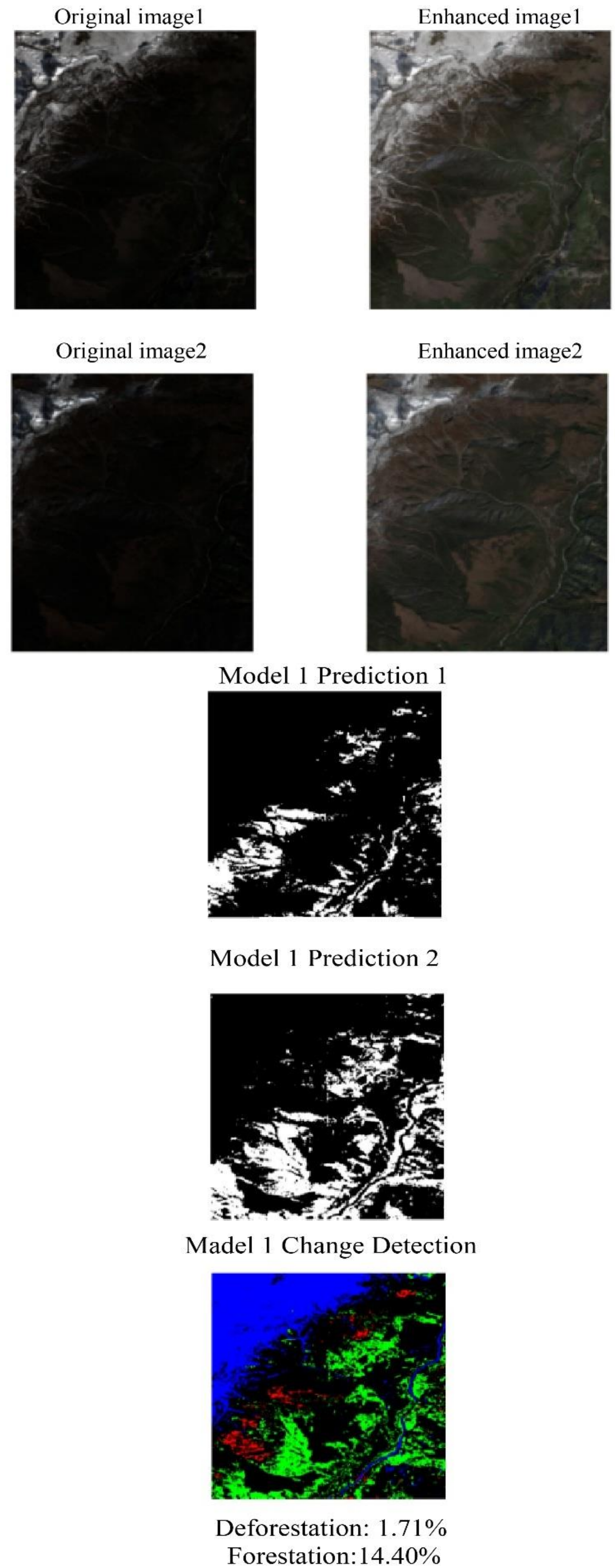



**Fig. 8 Snow detection and its elimination while calculating the percentage of forest change**

Table 2 presents the performance metrics evaluated for standard U-net, Residual U-net, and AResU-Net. As indicated in Table 2, AResU-Net outperforms the other two models.

**Table 2. Validation results of U-Net, residual U-Net and proposed hybrid model**

| Model | Evaluation Metrics | | | |
|---|---|---|---|---|
| | Accuracy | Precision | Recall | F1 Score |
| AResU-Net | 0.964782 | 0.946866 | 0.968095 | 0.957363 |
| Residual U-net | 0.9352 | 0.905 | 0.944 | 0.923 |
| Unet | 0.9332 | 0.8982 | 0.943 | 0.920 |

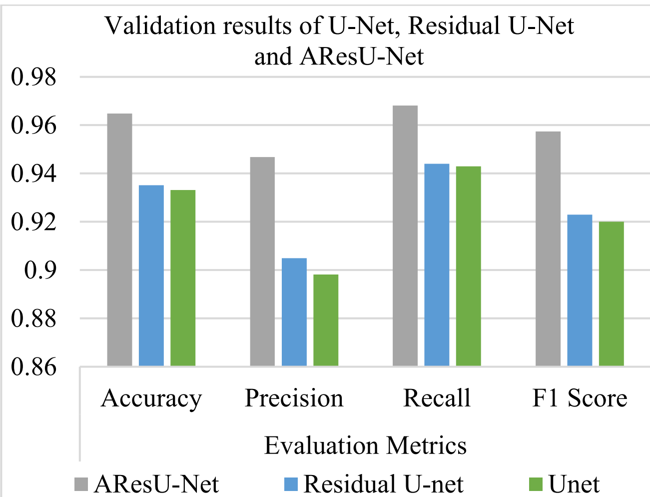



**Fig. 9 Validation results of U-Net, residual U-Net and AResU-net**

The results presented in Table 2 are represented in graphical form to compare each evaluation metric, which is shown in Figure 9.

The figure shows that the proposed model demonstrated superior performance to the U-Net and Residual U-net models. Figure 10 shows the screenshots of the evaluation metrics obtained for the models.

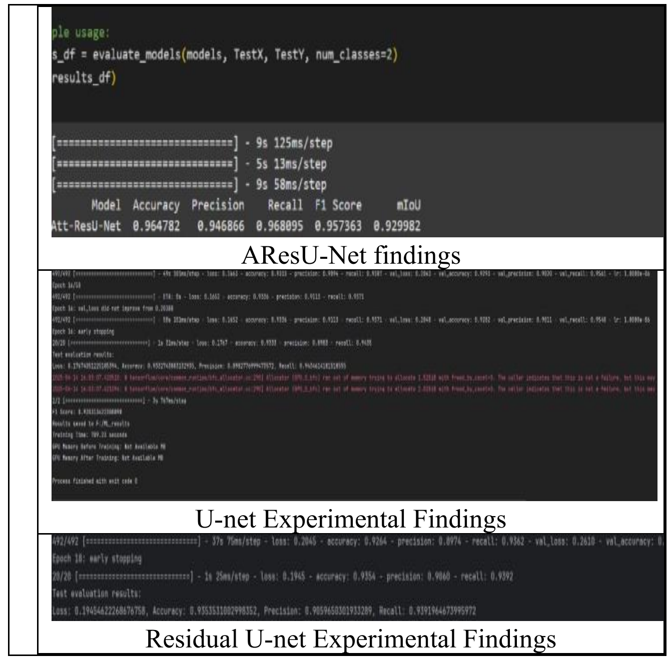



**Fig. 10 Screenshots of the evaluation metrics obtained for the models**

**Table 3. Comparison of the AResU-Net performance over existing state-of-the-art models**

| Paper Ref. No. | Methodology used | Evaluation metrics | | | |
|---|---|---|---|---|---|
| | | Precision | Recall | F1-score | Accuracy |
| [6] | Attention U-net with sentinel2 imagery for deforestation detection<br>Training Amazon - Atlantic Forest and testing Atlantic Forest - Amazon | 0.9222 | 0.8829 | 0.9021 | - |
| | U-net | 0.9169 | 0.8847 | 0.9005 | - |
| [5] | TransU-Net++: Deforestation mapping using attention gated U-net. | 0.7626 | 0.9087 | 0.8310 | 0.8821 |
| [10] | Attention-based semantic segmentation approach. HRNet+CBAM: Improvement over previous HRNet by combining with convolutional block attention module (abbreviated as CBAM) Using Landsat8 dataset | - | - | - | 0.9224 |
| [7] | Attention U-net demonstrated the best prediction for both F1 and IoU. | - | - | 0.90 | - |
| [11] | Channel Attention Mechanism (CAM). | 0.9491 | 0.8016 | 0.8691 | - |
| | Spatial Attention Mechanism (SAM) | 0.9428 | 0.8080 | 0.8702 | - |
| | SAM + CAM<br>Amazon forest area Landsat dataset | 0.9399 | 0.8162 | 0.8736 | - |
| [9] | ForestViT-Vision transformer approach using Amazon forest images | 0.80 | 0.94 | - | - |
| [24] | DeepLabv3+ using Landsat OLI-8 | 0.7176 | 0.7229 | 0.7180 | - |

| | | | | | |
|---|---|---|---|---|---|
| [14] | Results obtained using Res-Unet | 0.82.3 | 0.74 | 0.78 | - |
| AResU-Net Proposed model | Forest change detection semantic segmentation using attention-based Residual U-net with snow mask detection. | 0.946866 | 0.968095 | 0.957363 | 0.964782 |

Table 3 outlines the comparison of the AResU-Net model with existing models. The study by K. Kalinaki, O. A. Malik, and D. T. Ching Lai [2] introduced the Attention residual U-net for forest change detection. Using a Dataset of 924 Sentinel2 images, the study [2] achieved a remarkable Mean Intersection over Union(IOU) of -0.9330. The proposed AResU-Net model also demonstrated a high Mean Intersection over Union (mIOU)- 0.929982. Findings by David John et al. [6] highlighted the improvement in accuracy using attention over the U-net model. The presented findings of AResU-Net align with those of K. Kalinaki et al. [2] and David John et al. [6]. The results of the evaluation of existing recent findings [5, 7, 10, 11, 14, 24], also presented in Table 3, vary in dataset usage, methodologies followed for the implementation and evaluation metrics compared to the proposed method. AResU-Net performance is improved over U-Net and residual U-net due to incorporating attention gates and residual blocks for semantic segmentation. This improvement can inspire further research and development in this area.

## 5. Conclusion

This study introduces AResU-Net, a sophisticated hybrid attention-based U-Net methodology aimed at achieving semantic segmentation to detect changes in forest cover. The model can find the percentage of deforestation and afforestation by pixel-by-pixel comparison of a given pair of temporal sentinel2 images. Importantly, the model is designed to handle seasonal variations, especially during winter, by calculating snow/ice cover using NDVI, NDSI spectral indices and binary dilation. The calculated snow mask is eliminated from the temporal binary masks before calculating the percentage of forest change. Removing snow cover from the binary mask significantly reduces the risk of misclassifying snow cover as deforestation, thereby enhancing the model's accuracy. The model performance is validated using unseen sentinel2 images acquired using Google Earth Engine. Also, the model is implemented to handle input images of any size. The model is validated by acquiring images from the Nepal region, including snow. The performance of the model is compared with existing findings. Visual results obtained and evaluation metrics show similarity, which underscores the performance of the proposed model.

Based on the proposed work, the following future work is suggested: i) evaluation of the processing time of each model and presentation of its comparative analysis. ii) Use a pre-trained model to check its impact on processing time. iii) Use an approach to eliminate haze clouds to improve change detection. iii) Improve NDSI accuracy by including the impact of topography data.

## References


[1] Kostiantyn Isaienkov et al., "Deep Learning for Regular Change Detection in Ukrainian Forest Ecosystem with Sentinel-2," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 364-376, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] Kassim Kalinaki, Owais Ahmed Malik, and Daphne Teck Ching Lai, "FCD-Attresu-Net: An Improved Forest Change Detection in Sentinel-2 Satellite Images Using Attention Residual U-Net," *International Journal of Applied Earth Observation and Geoinformation*, vol. 122, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Flavie Pelletier et al., "Inter- and Intra-Year Forest Change Detection and Monitoring of Aboveground Biomass Dynamics Using Sentinel-2 and Landsat," *Remote Sensing of Environment*, vol. 301, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Ana María Pacheco-Pascagaza et al., "Near Real-Time Change Detection System Using Sentinel-2 and Machine Learning: A Test for Mexican and Colombian Forests," *Remote Sensing*, vol. 14, no. 3, pp. 1-21, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Ali Jamali et al., "TransU-Net++: Rethinking Attention Gated Trans U-Net for Deforestation Mapping," *International Journal of Applied Earth Observation and Geoinformation*, vol. 120, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] David John, and Ce Zhang, "An Attention-Based U-Net for Detecting Deforestation Within Satellite Sensor Imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 107, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Dongoo Lee, and Yeonju Choi, "A Learning Strategy for Amazon Deforestation Estimations Using Multi-Modal Satellite Imagery," *Remote Sensing*, vol. 15, no. 21, pp. 1-20, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] John Burns Kilbride et al., "A Near Real-Time Mapping of Tropical Forest Disturbance Using SAR and Semantic Segmentation in Google Earth Engine," *Remote Sensing*, vol. 15, no. 21, pp. 1-20, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Maria Kaselimi et al., "A Vision Transformer Model for Convolution-Free Multilabel Classification of Satellite Imagery in Deforestation Monitoring," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 7, pp. 3299-3307, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] See Ven Lim et al., "Attention-Based Semantic Segmentation Networks for Forest Applications," *Forests*, vol. 14, no. 12, pp. 1-22, 2023.

[CrossRef] [Google Scholar] [Publisher Link]

[11] P. Tovar, M.O. Adarme, and R.Q. Feitosa, "Deforestation Detection in the Amazon Rainforest with Spatial and Channel Attention Mechanisms," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43-B3-2021, pp. 851-858, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] K. Karaman, V. Sainte Fare Garnot, and J.D. Wegner, "Deforestation Detection in the Amazon with Sentinel-1 Sar Image Time Series," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 10-1/W1-2023, pp. 835-842, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Imran Md Jelas et al., "Deforestation Detection Using Deep Learning-Based Semantic Segmentation Techniques: A Systematic Review," *Frontiers in Forests and Global Change*, vol. 7, pp. 1-24, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Daliana Lobo Torres et al., "Deforestation Detection with Fully Convolutional Networks in the Amazon Forest from Landsat-8 and Sentinel-2 Images," *Remote Sensing*, vol. 13, no. 24, pp. 1-20, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[15] P.J. Soto et al., "Deforestation Detection with Weak Supervised Convolutional Neural Networks in Tropical Biomes," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43-B3-2022, pp. 713-719, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Alexandru-Toma Andrei, and Ovidiu Grigore, "Low-Cost Optimized U-Net Model with GMM Automatic Labeling Used in Forest Semantic Segmentation," *Sensors*, vol. 23, no. 21, pp. 1-20, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] Bruno Menini Matosak et al., "Mapping Deforestation in Cerrado Based on Hybrid Deep Learning Architecture and Medium Spatial Resolution Satellite Time Series," *Remote Sensing*, vol. 14, no. 1, pp. 1-22, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Zhipan Wang et al., "Repddnet: A Fast and Accurate Deforestation Detection Model with High-Resolution Remote Sensing Image," *International Journal of Digital Earth*, vol. 16, no. 1, pp. 2013-2033, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Abdul Haseeb Azizi et al., "Machine Learning-Based Estimation of Fractional Snow Cover in the Hindukush Mountains Using MODIS And Landsat Data," *Journal of Hydrology*, vol. 638, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Charlotte Poussin, Pascal Peduzzi, and Gregory Giuliani, "Snow Observation from Space: An Approach to Improving Snow Cover Detection Using Four Decades of Landsat and Sentinel-2 Imageries Across Switzerland," *Science of Remote Sensing*, vol. 11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[21] Yunlong Wang, and Jianshun Wang, "Monitoring Snow Cover in Typical Forested Areas Using a Multi-Spectral Feature Fusion Approach," *Atmosphere*, vol. 15, no. 4, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Gongxue Wang et al., "Characterization of NDSI Variation: Implications for Snow Cover Mapping," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[23] Rui He et al., "Effective Improvement of the Accuracy of Snow Cover Discrimination Using a Random Forests Algorithm Considering Multiple Factors: A Case Study of the Three-Rivers Headwater Region, Tibet Plateau," *Remote Sensing*, vol. 15, no. 19, pp. 1-21, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[24] Renan Bides de Andrade, Guilherme Lucio Abelha Mota, and Gilson Alexandre Ostwald Pedro da Costa, "Deforestation Detection in the Amazon Using DeepLabv3+ Semantic Segmentation Model Variants," *Remote Sensing*, vol. 14, no. 19, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]