

Imperial College London
Department of Earth Science and Engineering
MSc in Environmental Data Science and Machine Learning

Independent Research Project
Final Report

Monitoring Coastal Infrastructures in the Maldives using Remote Sensing and Machine Learning

by
Yibing Chen

Email: yibing.chen23@imperial.ac.uk

GitHub username: edsml-yc4523

Repository: <https://github.com/ese-msc-2023/irp-yc4523>

Supervisors:

Dr. Yves Plancherel

Ms. Myriam Prasow-Emond

August 2024

Abstract

As a Small Island Developing State (SIDS), the Maldives is particularly vulnerable to environmental challenges. Its low-lying topography makes the islands highly susceptible to shape alterations from sea level rise and coastal infrastructure development. This project employs remote sensing techniques and image segmentation models to monitor and analyze the development of coastal infrastructure in the Maldives. By utilizing Sentinel-2 satellite imagery, models such as U-Net, U-Net++, and DeepLabV3+ are applied to identify and track the evolution of harbours, jetties, and resorts. A multispectral data fusion approach, which combines RGB, NIR, NDWI, and NDVI bands, enhances the model's ability to distinguish between similar coastal features in a medium-resolution environment. The U-Net++ model achieved a mean Intersection over Union (mIoU) of 0.6520 and a Macro F1 score of 0.7674, showing superior performance in complex scenes, despite challenges related to class imbalance and image resolution. Change detection was also performed to analyze the temporal evolution of these infrastructures. The results are visualized using interactive maps created with Folium, providing researchers with a powerful tool to assess the impact of coastal infrastructure development on island morphology. This project enhances understanding of coastal infrastructure dynamics, providing tools that support the sustainable development of the Maldives.

1 Introduction

As a Small Island Developing State (SIDS), the Maldives faces unique challenges due to its geographical vulnerabilities. Most of its islands are low-lying, with an average elevation of around 1 meter above sea level, making them highly vulnerable to sea level rise [1]. The vulnerability is compounded by the country's heavy reliance on tourism, which is the largest economic pillar of the Maldives. Tourism consistently contributes 25%-30% to the Maldives' GDP and serves as the primary source of foreign exchange income and employment [2]. The Maldives is famous for its beautiful beaches and coral reefs, attracting many international tourists. Coastal infrastructures such as harbours, jetties, and resorts play a vital role in supporting and maintaining this industry. This economic development model has forced the country to engage in large-scale land reclamation projects and coastal infrastructure development [3].

However, the development and maintenance of these man-made coastal infrastructures profoundly affect the island's geomorphology and ecosystems. Existing research has explored how artificial structures like harbours, jetties, and resorts impact natural systems, particularly by altering sediment transport, salinity, and hydrodynamic processes [4]. These changes can lead to significant geomorphological shifts, such as erosion and sedimentation [5]. Despite the significance of these issues, no research has yet focused on how these man-made structures in the Maldives have changed over time and space and how such changes have influenced the shape of the islands.

In this context, remote sensing technology provides an important tool for addressing these environmental challenges. The Sentinel-2 satellite provides medium-resolution imagery regularly, making long-term monitoring of coastal areas possible [6]. By combining the

imagery with image segmentation models like U-Net [7], U-net++ [8] and DeepLabV3+ [9], we can achieve precise identification of coastal infrastructure. This project builds on these technologies to develop a tool that can automatically detect the spatial and temporal evolution of coastal infrastructures in the Maldives. By identifying the status of these structures, such as existent, non-existent and under construction, researchers can create a detailed timeline of coastal infrastructure development. This timeline can serve as a reference for studies exploring correlations between the status of these structures and broader environmental changes. For example, researchers could use this data in conjunction with other environmental data, such as sediment transport or coastal erosion, to investigate how these man-made structures may influence island morphology. By providing this foundational data, the project thus contributes to a deeper understanding of the complex interactions between human development and the environment, supporting more comprehensive studies on island morphology under the pressures of climate change and human development.

2 Methodology

2.1 Data Collection and Preprocessing

2.1.1 Data Acquisition

This project utilizes data obtained from Sentinel-2 satellite remote sensing images. The Sentinel-2 satellite is part of an Earth observation program initiated by the European Commission and the European Space Agency (ESA) to provide medium-resolution Earth observation data [10]. The Sentinel-2 satellite includes two identical satellites (Sentinel-2A and Sentinel-2B) that can offer global multispectral images with a spatial resolution of 10 to 60 meters and a repeat coverage capability every 5 days [11]. This makes it a powerful tool for studying coastal processes. The satellite data and images are available free of charge through the program.

We utilized the Python API of Google Earth Engine (GEE) to acquire 276 Sentinel-2 remote sensing images with a spatial resolution of up to 10 meters per pixel in GeoTIFF format, covering areas such as harbours, jetties, and resorts. The downloaded remote sensing images include RGB (red, green, blue) bands, NIR (near-infrared) bands, as well as NDVI (Normalized Difference Vegetation Index) and NDWI (Normalized Difference Water Index) bands, which were derived through calculation. The specific uses of the four bands and indices for identifying coastal infrastructure are shown in Table 1.

Table 1: Sentinel-2 Spectral Bands and Indices for Identifying Coastal Infrastructure

Band	Effect
RGB band (B2, B3, B4)	Closest to human vision. Highlights buildings and man-made structures by outlining their shapes.

Continued

Band	Effect
NIR band (B8)	Absorbed by water, reflected by buildings and bare ground. Helps distinguish water from man-made structures.
NDWI (Derived Index)	Highlights water features, making them distinct from other landforms like vegetation and buildings. Useful for identifying water-connected infrastructure.
NDVI (Derived Index)	Differentiates green areas from other landforms. Though mainly for vegetation, it also helps in identifying structures by contrasting them with vegetation.

Based on the statistical analysis, the median aspect ratio of images in the dataset is 147x112. To preserve image quality and avoid distortion, all images are transformed into JPG format and uniformly resized to 128x128 pixels.

2.1.2 Data Annotation

Since the images of these four bands and indices are consistent in geographical location and size, this project selected the NDWI band for polygon annotation using Roboflow, an online image annotation platform. After the annotation was completed, the polygon information was exported and converted into a mask image in PNG format. Figure 1 shows the images of four bands and indices along with the corresponding mask.

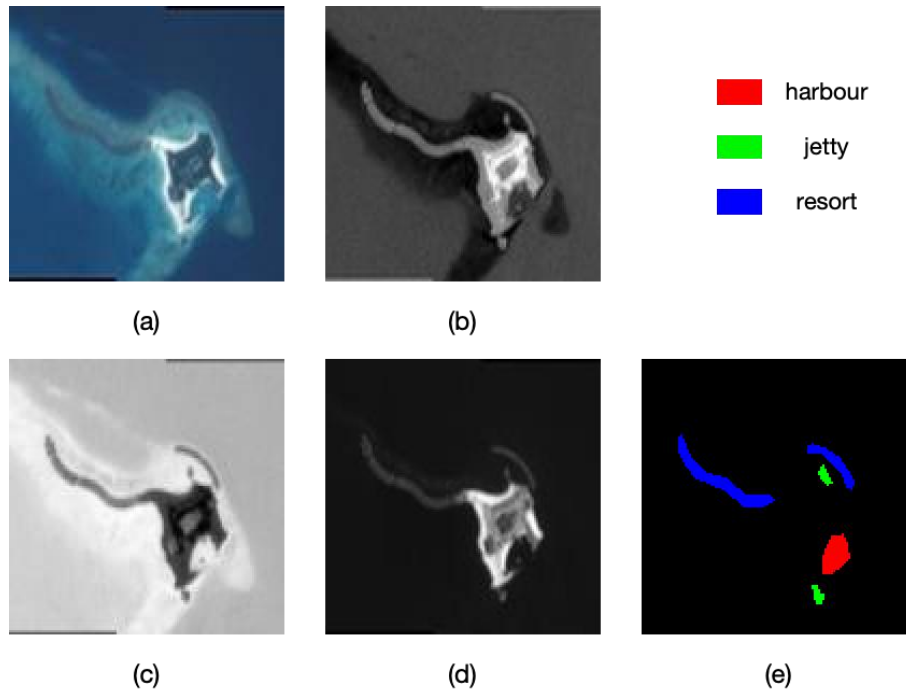


Figure 1: The location of the images is near Malé, the capital of Maldives, shown using different spectral bands and a corresponding mask image. (a) RGB band, (b) NIR band, (c) NDWI, (d)

NDVI, and (e) the mask image generated from annotations. In the mask, red represents the harbour, green represents the jetty, and blue represents the resort.

2.1.3 Data Preparation

This project combines RGB, NIR, NDWI and NDVI images to create a six-channel “.npy” format image file. This multispectral data fusion aims to merge information from different spectral bands to improve the image segmentation performance of the model. The dataset comprises 276 images, which were randomly divided into training, validation, and test sets in a ratio of 7:2:1. Given the small size of the dataset, the training set was augmented to reduce the risk of overfitting and enhance the generalization ability of the model. The specific image augmentation strategies include rotation, horizontal flipping, and vertical flipping, which increased the number of training samples by five times [12]. Additionally, the images underwent center cropping, where each image was cropped to 116x116 pixels and then resized back to 128x128 pixels, to focus the model’s attention on the most relevant features. The normalization process was also applied to the entire dataset. By calculating the mean and standard deviation of the pixel values, the image data for each channel was normalized to a distribution with a mean close to 0 and a standard deviation of 1 before training. This process ensures consistency in the model input data and enhances the stability of the training process.

2.2 Model Selection

This project utilizes a combination of different image segmentation models, including U-Net combined with ResNet-34, U-Net combined with ResNet-50, U-Net++ combined with ResNet-50, and DeepLabV3+ combined with ResNet-50. These models are used to achieve accurate segmentation of remote sensing images. Among them, the U-Net architecture is used as a baseline for comparison with other more complex models.

ResNet series is a deep convolutional neural network known for its ability to deal with the gradient vanishing problem during the training of deep neural networks. ResNet-34 and ResNet-50 are key versions in the ResNet series, with 34 and 50 convolutional layers respectively [13]. These models are used as backbone networks for feature extraction. They are pre-trained on the ImageNet dataset and improve the accuracy and convergence speed in the current task through transfer learning [14].

U-Net is a classic medical image segmentation architecture, particularly effective in tasks involving small samples. It uses a symmetrical encoder-decoder structure to restore details while maintaining global context information [7]. U-Net++ is an enhanced version of U-Net, which further improves the segmentation accuracy by introducing dense skip connections and deeper feature extraction [8]. DeepLabV3+ improves segmentation by combining atrous spatial pyramid pooling (ASPP) with a lightweight decoder. ASPP captures multi-scale information effectively, while the decoder refines segmentation, especially around object

boundaries. This combination allows the model to achieve more accurate results in complex scenes [9].

To adapt to the multispectral image data of this project, the number of input channels in the models was expanded from 3 to 6. This adjustment allows the models to process fused RGB, NIR, NDWI, and NDVI data. Additionally, dropout techniques were applied within the models to prevent overfitting by randomly deactivating a fraction of neurons during training, enhancing the model's generalization ability [15]. The output layer of each model was also adapted to handle the four-class segmentation task, namely background, harbour, jetty, and resort.

2.3 Model Training and Validation

2.3.1 Training Stages

The model training process is divided into three stages. This staged approach allows for gradual fine-tuning, beginning with foundational feature learning and progressing to more complex feature integration.

In the initial stage (Epoch 1-40), the shallow encoders are frozen, while the intermediate and deep layers are trained. This strategy allows the model to refine the more abstract features without altering the basic feature detectors. The batch size in this stage is set to 16, and the learning rate is gradually adjusted according to the cosine annealing strategy [16], which gradually reduces the learning rate as the number of epochs increases.

As the training progresses (Epoch 41-80), the intermediate layers are unfrozen to further optimize the intermediate feature representations. The batch size is reduced to 8 in this stage to accommodate the training needs of more layers, while the learning rate continues to be adjusted according to the cosine annealing strategy.

In the final stage (Epoch 81-120), all layers are unfrozen so that the entire network participates in the training. The batch size is further reduced to 4 to ensure the stability of the model as global parameters are updated.

At the end of each stage, the model is evaluated on the validation set to monitor the performance of the model and ensure its generalization ability across different tasks.

2.3.2 Loss Function

By analyzing the training set, we found that the distribution of pixels in different categories is extremely unbalanced. The background class occupies most pixel values in the dataset, while the pixel values of the main target classes, such as harbour, jetty and resort, are very small. Figure 2 shows the pixel counts for different classes in the log scale.

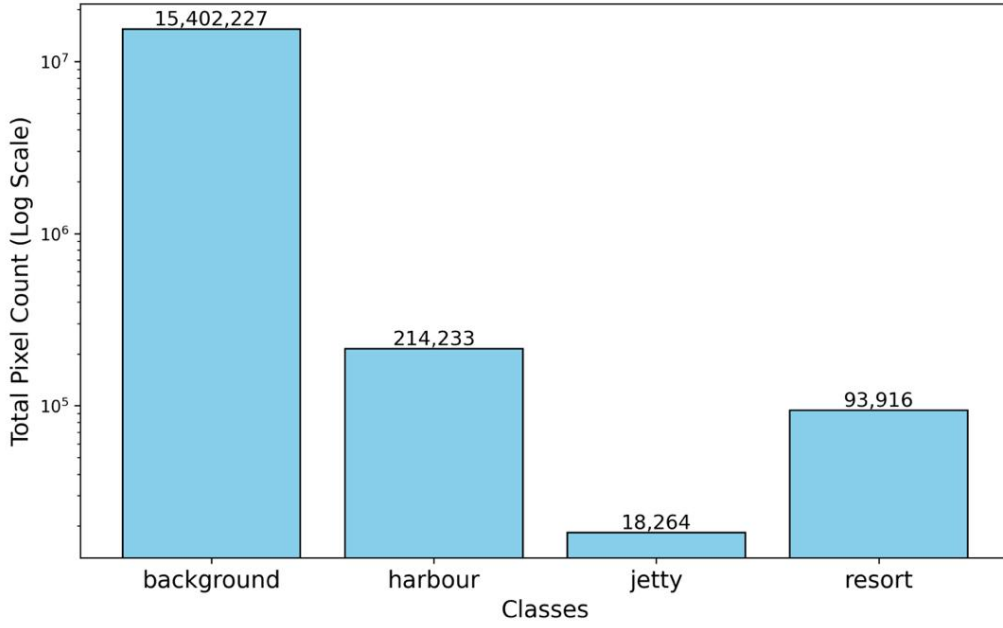


Figure 2: Pixel Counts for different classes in the log scale

A combined loss function that combines Dice loss and Focal loss was designed to reduce the impact of class imbalance on model training. Dice loss focuses on the overlap between the predicted area and the true area and has a better processing effect on small sample categories [17]. The calculation method is illustrated in Equation (1).

$$\text{Dice loss} = 1 - \frac{2 \sum_{i=1}^N p_i g_i + \text{smooth}}{\sum_{i=1}^N p_i + \sum_{i=1}^N g_i + \text{smooth}} \quad (1)$$

Where:

- p_i denotes the predicted probability of the i -th pixel by the model.
- g_i denotes the ground truth label of the i -th pixel.
- N is the total number of pixels.
- smooth is a small constant added to prevent division by zero.

Focal loss reduces the dominant influence of background class on the overall loss by decreasing the weight of easily classified samples and enhances the model's learning ability for small sample classes [18]. The calculation method is illustrated in Equation (2).

$$\text{Focal Loss} = -\alpha(1 - p_t)^{\gamma} \log(p_t) \quad (2)$$

where:

- p_t represents the model's predicted probability for the correct category.
- α is a balancing factor used to adjust the imbalance between positive and negative samples.

- γ is a modulation factor used to reduce the loss contribution of samples that are easy to classify, usually with a value of 2.

2.3.3 Optimizer and Learning Rate Scheduling

During the model training process, the AdamW optimizer was used to update the model parameters. This optimizer improves upon the traditional Adam optimizer by introducing weight decay, which can more effectively prevent the model from overfitting [19]. The initial learning rate was set to 1×10^{-4} , with a minimum learning rate of 1×10^{-6} , which is 1% of the initial rate. To adapt to the impact of different batch sizes on model training, an adaptive learning rate adjustment mechanism was adopted. It dynamically adjusts the learning rate according to the batch size to ensure stable training of the model.

To better manage the change in learning rate, a cosine annealing learning rate scheduler was applied [16]. During the initial stage of training, a warm-up period was implemented to prevent the negative impact of a high learning rate on model convergence. The learning rate gradually increased from a lower value to the initial learning rate. In subsequent stages, it decreased following the cosine curve and gradually approached the minimum value near the end of training. In this way, the instability caused by drastic fluctuations in the learning rate can be effectively avoided.

2.4 Model Testing

2.4.1 Mean Intersection over Union

Mean Intersection over Union (mIoU) measures the segmentation accuracy of the model by calculating the ratio of the overlap and union of the predicted area and the true area [20]. The calculation steps of mIoU are as follows:

For each class i , its intersection over union (IoU) is defined as (Equation (3)):

$$\text{IoU}_i = \frac{TP_i}{TP_i + FP_i + FN_i} \quad (3)$$

where:

- TP_i : The number of pixels correctly predicted as class i by the model.
- FP_i : The number of pixels incorrectly predicted as class i by the model.
- FN_i : The number of pixels that belong to class i but were predicted by the model as another class.

Mean Intersection over Union (mIoU) is calculated as the average of the IoU values for each class (Equation (4)):

$$mIoU = \frac{1}{N} \sum_{i=1}^N IoU_i \quad (4)$$

Where: N is the total number of classes. In this project, $N = 4$ (corresponding to background, harbour, jetty, and resort).

2.4.2 Macro F1 Score

Macro F1 Score plays an important role in unbalanced datasets to prevent the model from performing well only on the majority class (background) and ignoring the minority class (harbour, jetty and resort). It provides a more fair and comprehensive way to evaluate the actual segmentation ability of the model on all classes by calculating the harmonic mean of precision and recall [21]. This is especially critical in image segmentation tasks, where misclassification can lead to significant costs like erroneous identification of key features. The calculation steps of the macro F1 score are as follows:

For each class i , its F1 score is defined as (Equation (5)):

$$F1\ Score_i = 2 \times \frac{Precision_i \times Recall_i}{Precision_i + Recall_i} \quad (5)$$

where:

- $Precision_i$: represents the proportion of pixels predicted as positive for class i that are correctly predicted by the model. The Equation (5-1) shows the calculation steps:

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad (5-1)$$

- $Recall_i$: represents the proportion of pixels that are truly positive for class i and are correctly predicted by the model. The Equation (5-2) shows the calculation steps:

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (5-2)$$

Then, the average of F1 Scores across all classes is calculated to obtain the Mean F1 Score (Equation (6)), which serves as an evaluation metric for the overall segmentation performance of the model.

$$Macro\ F1\ Score = \frac{1}{N} \sum_{i=1}^N F1\ Score \quad (6)$$

Where: N is the total number of classes. In this project, $N = 4$ (corresponding to background, harbour, jetty, and resort).

2.5 Change Detection and Mapping

2.5.1 Image Inference Using the Trained Model

This project targets areas in the Maldives, such as Addu Atoll, which are experiencing rapid economic development and are particularly vulnerable to climate change. Sentinel-2 remote sensing images of harbours, jetties, and resorts in these regions were obtained using Google Earth Engine (GEE) over the past decade. For each specific area, the trained model was applied to images from different periods to generate the corresponding segmentation masks.

2.5.2 Mask Processing and Landmark Attribute Extraction

Morphological operations were applied to the generated masks to remove noise and isolated pixels, ensuring their clarity and accuracy.

Next, coastal infrastructure in the images was identified, and their attributes were extracted. Specifically, the geometric center (centroid) and area of each connected region were detected, corresponding to the location and size of the coastal infrastructure. The extracted information, including the detection date, type, centroid coordinates, and area, was systematically recorded in a data table to facilitate subsequent change detection and spatial analysis. If no coastal infrastructure was detected in a particular image, the corresponding null value was recorded in the data table.

To transform the detected infrastructure from image coordinates to geographic coordinates, a pixel-based geographic conversion method was used. Considering the geographic boundaries of the study area and the spatial resolution of the image, the latitude and longitude of each structure's centroid were calculated.

The processed data was organized and stored in a structured data file to ensure that the new detection results were saved together with the existing data for subsequent analysis.

2.5.3 Detection Analysis

Based on the extracted information, the status of each coastal infrastructure was determined and classified as follows:

- (1) For structures detected at each time node, their status was set to "existing."
- (2) For structures that were not detected, their status was set to "non-existent."
- (3) For structures that existed in the early stage but disappeared later, their status at the time node when they began to disappear was set to "demolished."
- (4) For structures that did not exist in the early stage but appeared later, their status at the time node when they were first detected was set to "newly constructed."

- (5) For structures that were not initially detected but later detected as being under construction, their status was set to "under construction" until the construction was completed, at which point their status was changed to "construction completed."
- (6) Additionally, for structures that existed in the early stage and showed regional expansion later, their status was set to "expansion began" at the beginning of the expansion and marked as "expansion completed" after the expansion.
- (7) Some structures may undergo continuous changes over a long period, such as gradual expansion or demolition. In these cases, the status of these structures will be marked as "continuously expanding" or "continuously demolishing," with the specific duration recorded in the data for more detailed trend analysis.

Occasionally, false "newly constructed" or "continuously expanding" annotations may appear due to errors in image quality or segmentation models. When this occurs, it may be necessary to confirm the accuracy of these annotations through post-verification to avoid the spread of misinformation.

2.5.4 Visualization

Folium is an interactive mapping tool that integrates seamlessly with common data processing libraries such as Pandas and GeoPandas. It enables users to convert geographic data into visually interactive maps. By leveraging Folium's extensive capabilities, users can overlay various types of data, customize map aesthetics, and create dynamic interactive features such as pop-ups, tooltips, and time sliders [22].

This tool will be used to display the detection results containing geographic locations and attributes of structures on an interactive map, enabling intuitive analysis of spatial distribution and temporal trends.

3 Results

3.1 Histogram of mIoU and Macro F1 Score

Figure 3 illustrates the performance of different models in terms of mIoU and mean F1 score. Among the models tested, U-Net++-ResNet-50 performed the best on the test set, with a mIoU of 0.6520 and a Macro F1 score of 0.7674. The Unet-ResNet-50 model followed, achieving a mIoU of 0.6294 and a Macro F1 score of 0.7427. DeeplabV3+-ResNet-50 recorded a mIoU of 0.6171 and a Macro F1 score of 0.7316. Lastly, the U-Net-ResNet-34 model demonstrated a mIoU of 0.6103 and a Macro F1 score of 0.7195. This comparison underscores the superior performance of U-Net++-ResNet50.

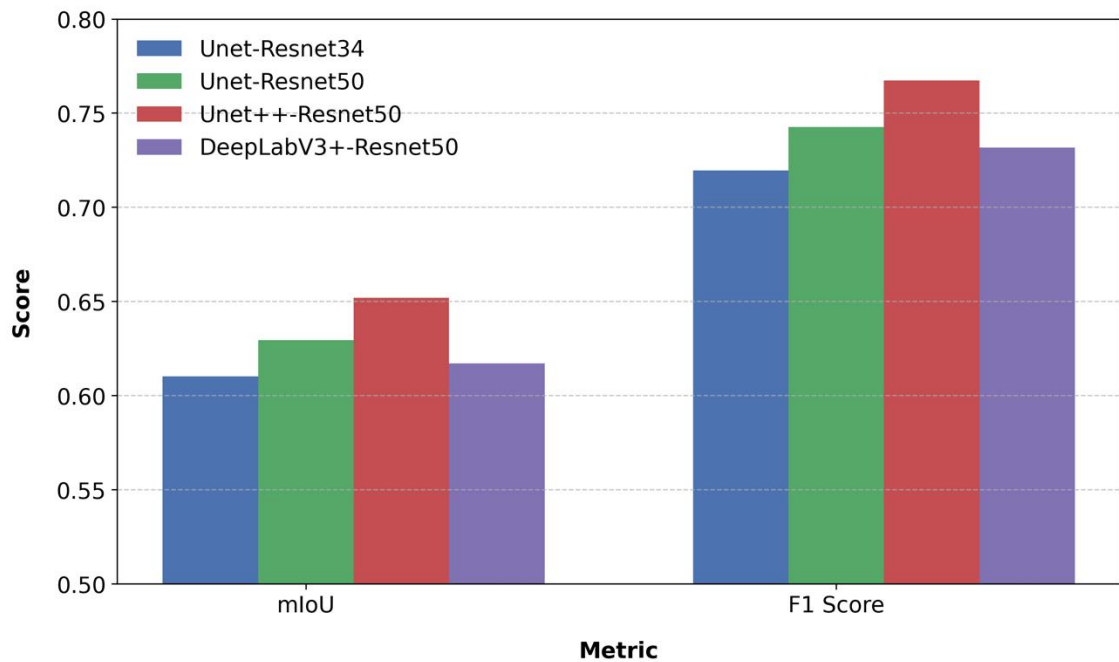


Figure 3: Comparison of mIoU and mean F1 Score across different models



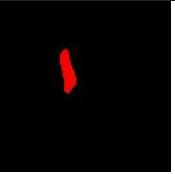
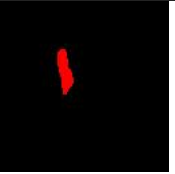

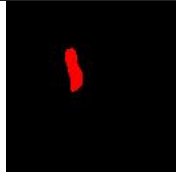

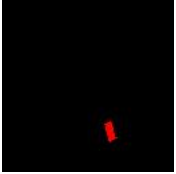
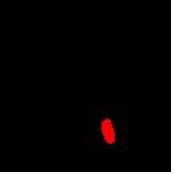
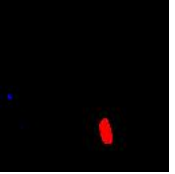
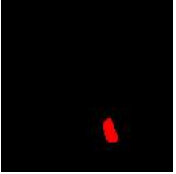


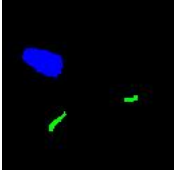
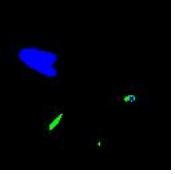
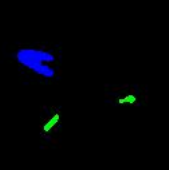
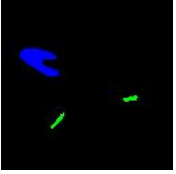
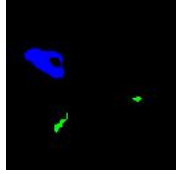

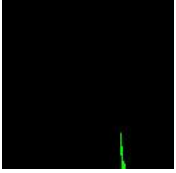
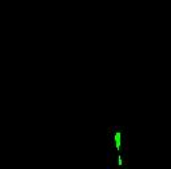
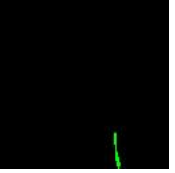
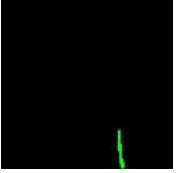
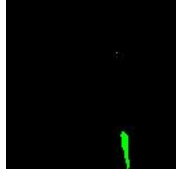
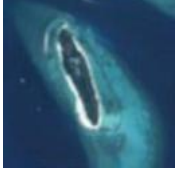
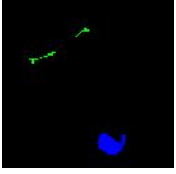
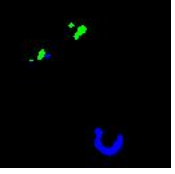
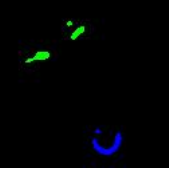
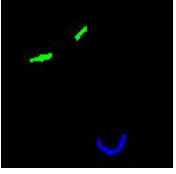
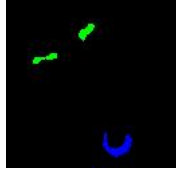

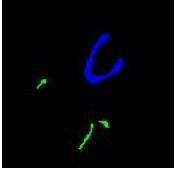
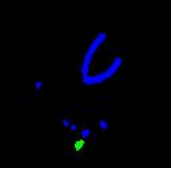

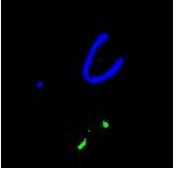
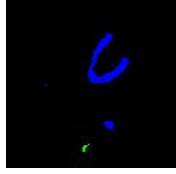
3.2 Comparison of segmentation results of each model

Table 2 compares the segmentation results from the models UNet-ResNet-34, UNet-ResNet-50, UNetPlusPlus-ResNet-50, and DeepLabV3Plus-ResNet-50, focusing on their accuracy in segmenting coastal infrastructure from remote sensing images. The UNet-ResNet-34 model shows moderate accuracy but struggles with complex scenes. The UNet-ResNet-50 model improves on this, handling details better in larger, less cluttered areas. The UNetPlusPlus-ResNet-50 model performs the best, accurately segmenting smaller and more complex features with clear boundaries. The DeepLabV3Plus-ResNet-50 model is effective but sometimes struggles with fine details and smaller structures, though it performs well in simpler scenes.

Table 2: A selection of input images, their corresponding reference masks, and the output masks predicted by various segmentation models (a-g). The reference mask represents the ground truth segmentation, while the model outputs illustrate the predicted segmentations based on the multi-channel input images.

	Image (RGB)	Reference mask	Model Output			
			<i>UNet- ResNet-34</i>	<i>UNet- ResNet-50</i>	<i>UNetPlusPlus- ResNet-50</i>	<i>DeepLabV3Plus- ResNet-50</i>
(a)						

continued

	Image (RGB)	Reference mask	Model Output			
			<i>UNet- ResNet-34</i>	<i>UNet- ResNet-50</i>	<i>UNetPlusPlus- ResNet-50</i>	<i>DeepLabV3Plus- ResNet-50</i>
(b)						
(c)						
(d)						
(e)						
(f)						
(g)						

3.3 IoU and F1 Score for Each Class Across Different Models

To further explore the segmentation performance of each model on individual classes, Figure 4 presents the Intersection over Union (IoU) and F1 scores for each class (Background, Harbour, Jetty, and Resort) across the four different models

All models performed very well in the Background class, with IoU and F1 scores close to 0.99. However, their performance differed in smaller classes like Jetty. The Unet++-ResNet50 model stood out by achieving the highest IoU (0.3657) and F1 (0.5355) scores for the Jetty class, showing its strength in handling smaller, more complex features. This model

also performed best in the Resort class, making it the most effective for these challenging classes.

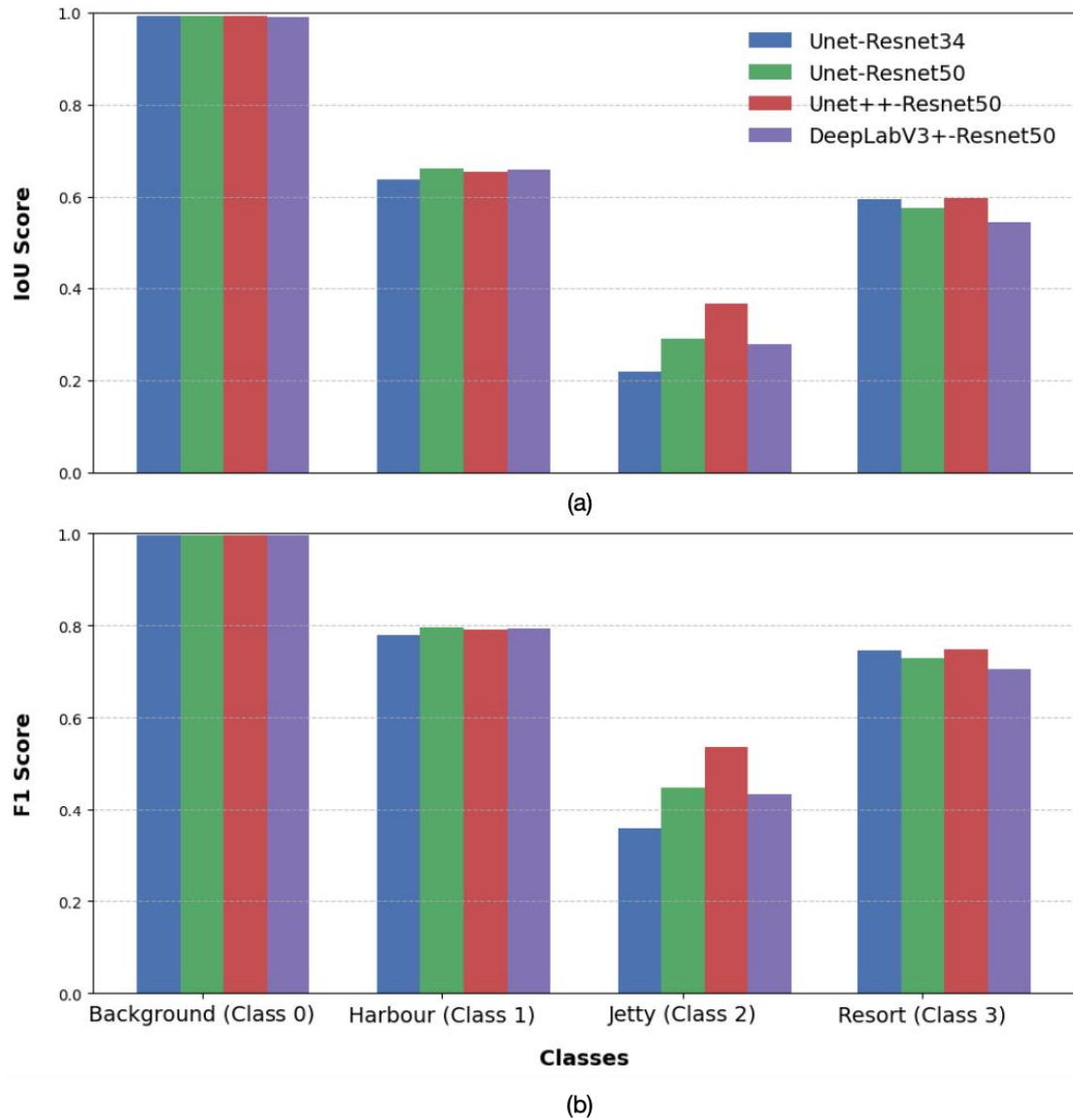


Figure 4: Comparison of IoU and F1 Scores across Different Models for Each Class. The top graph (a) shows the IoU scores for the Background, Harbour, Jetty, and Resort classes, while the bottom graph (b) presents the corresponding F1 scores.

3.4 Change Detection and Visualization

After performing image inference using the trained U-Net++-ResNet-50 model, the results of the change detection analysis were stored in a structured data file. This file contained detailed information about the detected changes in coastal infrastructure, including the status, latitude and longitude, and type of each structure over time. In the visualization, coastal infrastructure was represented by different shapes and colors to denote their types and statuses. This differentiation allowed users to quickly grasp the current state and historical changes of each structure. For example, regarding status, "Existing" infrastructure

was marked in grey, while "Non-existent" structures were represented in red. In terms of types, "Harbors" were depicted as filled circles, "Jetties" as half-filled circles, and "Resorts" as unfilled circles, making it easy to distinguish between them.

A dynamic timeline was implemented on the map, allowing users to interactively explore how coastal infrastructure evolved over time. The timeline slider provided a month-by-month view, illustrating the gradual changes in the region. Additionally, a custom legend was incorporated into the map, explaining the colour-coding and shapes used to represent different statuses and types. This legend ensured that the visualization was both informative and easy to interpret, allowing users to quickly understand the data presented.

Figure 5 shows an interactive map created using Folium to detect changes in coastal infrastructure within the Addu Atoll. It clearly illustrates the spatial and temporal dynamics of these structures. The interactive nature of the map, coupled with detailed data integration, provided a powerful tool for exploring and analyzing the region's changes over time.

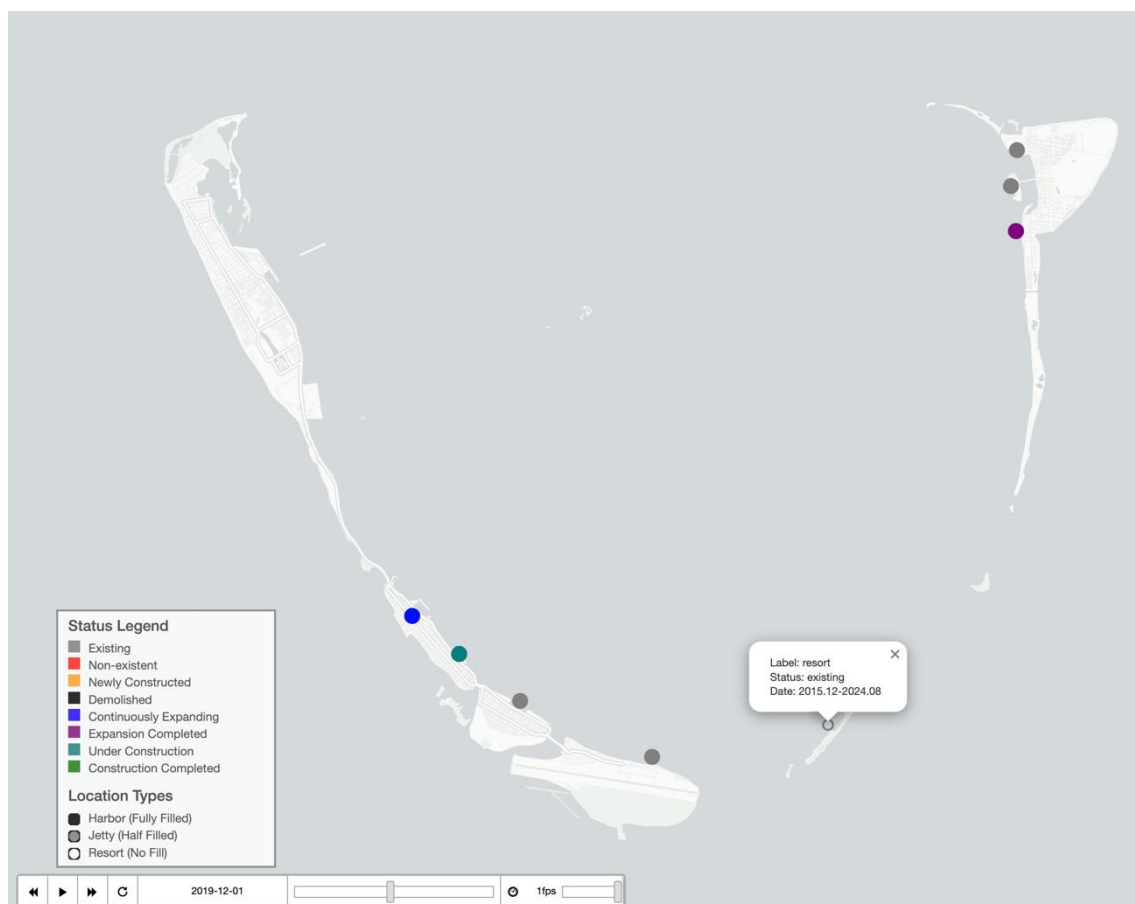


Figure 5: Interactive Folium map of the Addu Atoll, illustrating changes in coastal infrastructure and providing detailed status information for each location.

4 Discussion

4.1 Model performance

Among all the models used, U-Net++ combined with ResNet-50 showed the best results, achieving the highest mIoU and Macro F1 score (0.6520 and 0.7674, respectively). This model's strong performance is likely due to its deep feature extraction and dense skip connections. These skip connections help pass and combine information from different levels, which is especially useful in complex backgrounds involving structures like jetties and resorts. This reduces information loss and improves the model's ability to handle details and boundaries. In contrast, U-Net and DeepLabV3+ perform less effectively possibly because they lack enough context information processing in complex scenes, leading to errors when segmenting small, intricate objects like jetties. This is also reflected in the confusion matrix, where the classification error rate of U-Net and DeepLabV3+ models in the jetty is significantly higher than that of U-Net++.

4.2 Class imbalance

A significant challenge encountered in this project was the class imbalance. The background class dominated the dataset, while the primary classes of interest such as harbour, jetty, and resort are underrepresented. In particular, the jetty not only occupied the fewest pixels but was also difficult to identify due to its small size. This imbalance made it harder for the models to learn and accurately predict the smaller and less frequent classes.

While methodology includes strategies like combining Dice loss with Focal loss and applying data augmentation techniques such as center cropping, these were not sufficient to completely overcome the problem of class imbalance. The combined loss function aims to address the disproportionate influence of the background class by focusing the model's attention on smaller objects, while center cropping highlights more critical areas of the image by reducing the number of background pixels. However, despite these efforts, the models exhibited varied performance across different classes. For example, in the U-Net++ combined with ResNet-50 model, the mIoU values for the background, harbour, jetty, and resort are 0.9920, 0.6537, 0.3661, and 0.5966, respectively. This shows that even with advanced techniques to address class imbalance, identifying small objects like the jetty is still challenging. These objects have minimal representation in the dataset, which makes model training more difficult.

4.3 Image resolution

Another challenge was the medium resolution of Sentinel-2 remote sensing images. Although this resolution is sufficient for capturing broad features, it makes identifying smaller objects like jetties more difficult. The medium resolution can make the model miss fine

details, which are crucial for accurately distinguishing between similar classes such as jetties and resorts. Even to the human eye, it can sometimes be difficult to distinguish features between them. Additionally, the background often includes islands, coral reefs, and other features, making the scene more cluttered. In Sentinel-2 remote sensing images, these elements can distract the model and cause mistakes when trying to accurately identify resorts or jetties. This complexity also introduces a certain degree of inaccuracy during the image annotation process, further complicating the task of precise feature identification.

Despite the limitations of medium resolution, fusing multispectral data (combining RGB, NIR, NDWI, and NDVI bands into a six-channel image format) helps address this challenge. This approach compensates for the medium resolution by providing diverse spectral information, which enhances the model's ability to detect subtle differences. Each spectral band offers complementary details: for example, the NIR band is particularly effective at distinguishing water from man-made structures, while the NDWI band highlights water features, aiding in the accurate segmentation of coastal infrastructure. By using multispectral fusion, the model can leverage various aspects of the data that might not be as apparent in standard RGB imagery.

4.4 Application of coastal infrastructure monitoring

As a Small Island Developing State (SIDS) facing the challenges of rising sea levels and other impacts of climate change, the Maldives must balance economic development with environmental protection. The interactive map created using Folium can visually display the spatial distribution and temporal evolution of coastal infrastructure.

For example, on Maradhoo Island in Addu Atoll, the harbour has experienced multiple phases of construction, demolition, and reconstruction from 2015 to 2024: From December 2015 to January 2017, the harbour was "non-existent"; it was then "newly constructed" in February 2017 but was "demolished" in March 2017. From March 2017 to February 2019, the harbour was again "non-existent" until it was rebuilt in March 2019, briefly "existing" before being "demolished" again in April 2019. From April to September 2019, the harbour remained "non-existent" until construction resumed in November 2019 and was "completed" in February 2022, remaining "existing" through August 2024.

This sequence of changes reflects the interaction between infrastructure development and environmental factors. On the one hand, rising sea levels can cause coastal erosion, and extreme weather events may endanger the safety and stability of infrastructure. These environmental factors have led the Maldives to repeatedly adjust and replan its infrastructure, taking measures such as demolition and reconstruction. On the other hand, infrastructure itself can also affect island morphology, because development projects such as harbours, jetties or resorts can change natural sediment flows, potentially accelerate coastal erosion and may increase erosion in other areas. Monitoring these changes allows researchers to analyze how infrastructure projects affect the physical shape and stability of the island.

Therefore, the interactive map created with Folium offers a platform to visualize the spatial and temporal distribution of infrastructure changes. By providing a detailed visualization of the status of these structures, the tool enables researchers to analyze potential correlations between infrastructure projects and environmental factors. This analysis offers valuable insights into how past and ongoing projects have influenced or altered the evolution of island morphology. These insights, in turn, help researchers assess the long-term sustainability of infrastructure projects and develop strategies to mitigate negative environmental impacts.

5 Conclusion

This project successfully applied image segmentation models to monitor and analyze the evolution of coastal infrastructure in the Maldives, focusing on harbours, jetties, and resorts. This work addresses a critical gap in the identification and monitoring of coastal infrastructure. By leveraging multispectral data fusion, the models effectively processed Sentinel-2 remote sensing images, capturing a more comprehensive view of these infrastructures. Despite challenges posed by class imbalance and the medium resolution of Sentinel-2 images, particularly in accurately identifying smaller structures like jetties, the use of techniques such as data augmentation and loss function customization mitigated some of these difficulties. Among the models tested, the U-Net++ combined with the ResNet-50 model proved the most effective, particularly in handling smaller and more intricate features like jetties.

The creation of an interactive map using Folium served as a platform for visualizing the spatial and temporal changes in coastal infrastructure. It also provided valuable information for researchers exploring the dynamic interactions between human development and environmental factors, particularly in understanding how infrastructure projects have influenced island morphology. The map thus contributes to a deeper understanding of the sustainable management of coastal areas in the face of climate change.

Reference

- [1] V. K. E. Duvat. Human-driven atoll island expansion in the Maldives. *Anthropocene*, 32:100265, 2020.
- [2] U.S. Department of State. 2023 Investment Climate Statements: Maldives. Retrieved from <https://www.state.gov/reports/2023-investment-climate-statements/maldives/>, 2023.
- [3] M. Aslam and P. S. Kench. Reef island dynamics and mechanisms of change in Huvadhoo Atoll, Republic of Maldives, Indian Ocean. *Anthropocene*, 18:57–68, 2017.
- [4] F. Bulleri and M. G. Chapman. The introduction of coastal infrastructure as a driver of change in marine environments. *J. Appl. Ecol.*, 47:26–35, 2010.
- [5] K. F. Nordstrom. Living with shore protection structures: a review. *Estuar. Coast. Shelf Sci.*, 150:11–21, 2014.

- [6] K. D. Splinter, M. D. Harley, and I. L. Turner. Remote sensing is changing our view of the coast: Insights from 40 years of monitoring at Narrabeen-Collaroy, Australia. *Remote Sens.*, 10(10):1744, 2018.
- [7] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III*, pages 234–241. Springer International Publishing, 2015.
- [8] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings*, pages 3–11, 2018.
- [9] L. C. Chen, Y. Zhu, G. Papandreou, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 801–818, 2018.
- [10] S. Jutz and M. P. Milagro-Perez. Copernicus: The European Earth Observation Programme. *Rev. De Teledetección*, 56–XI, 2020.
- [11] R. S. Inteti, V. R. Mandla, J. R. Peddada, and N. Ramesh. Analysis of rice crop phenology using Sentinel-1 and Sentinel-2 satellite data. In *Advances in Geotechnical and Transportation Engineering, Lecture Notes in Civil Engineering*, volume 71, pages 257–266, 2020.
- [12] C. Shorten and T. M. Khoshgoftaar. A survey on image data augmentation for deep learning. *J. Big Data*, 6:60, 2019.
- [13] P. Nagpal, S. Bhinge, and A. Shitole. A comparative analysis of ResNet architectures. In *2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, pages 1–8. IEEE, 2022.
- [14] M. Zou and Y. Zhong. Transfer learning for classification of optical satellite image. *Sensing and Imaging*, 19, 2018.
- [15] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15:1929–1958, 2014.
- [16] I. Loshchilov and F. Hutter. SGDR: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.
- [17] F. Milletari, N. Navab, and S. A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, 2016.

- [18] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2980–2988, 2017.
- [19] I. Loshchilov and F. Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- [20] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*, 2017.
- [21] J. Opitz and S. Burst. Macro f1 and macro f1. *arXiv preprint arXiv:1911.03347*, 2019.
- [22] Folium: Python Data, Leaflet.js Maps. (n.d.). Retrieved from <https://python-visualization.github.io/folium/latest/>