



## Article

# LinkNet-Spectral-Spatial-Temporal Transformer Based on Few-Shot Learning for Mangrove Loss Detection with Small Dataset

Ilham Adi Panuntun <sup>1</sup>, Ilham Jamaluddin <sup>2</sup>, Ying-Nong Chen <sup>1,2,\*</sup>, Shiou-Nu Lai <sup>3</sup> and Kuo-Chin Fan <sup>2</sup>

<sup>1</sup> Center for Space and Remote-Sensing Research, National Central University, No. 300, Jhongda Rd., Jhongli Dist., Taoyuan City 32001, Taiwan; ilhamadipanuntun@g.ncu.edu.tw

<sup>2</sup> Department of Computer Science and Information Engineering, National Central University, No. 300, Jhongda Rd., Jhongli Dist., Taoyuan City 32001, Taiwan; ilhamjamaluddin@g.ncu.edu.tw (I.J.); kcfan@csie.ncu.edu.tw (K.-C.F.)

<sup>3</sup> Department of Business Administration, Hsing Wu University, No. 101, Sec.1, Fenliao Rd., LinKou Dist., New Taipei City 244012, Taiwan; 091046@mail.hwu.edu.tw

\* Correspondence: yingnong1218@csrsr.ncu.edu.tw

**Abstract:** Mangroves grow in intertidal zones in tropical and subtropical regions, offering numerous advantages to humans and ecosystems. Mangrove monitoring is one of the important tasks to understand the current status of mangrove forests regarding their loss issues, including deforestation and degradation. Currently, satellite imagery is widely employed to monitor mangrove ecosystems. Sentinel-2 is an optical satellite imagery whose data are available for free, and which provides satellite imagery at a 5-day temporal resolution. Analyzing satellite images before and after loss can enhance our ability to detect mangrove loss. This paper introduces a LSST-Former model that considers the situation before and after mangrove loss to categorize non-mangrove areas, intact mangroves, and mangrove loss categories using Sentinel-2 images for a limited number of labels. The LSST-Former model was developed by integrating a fully convolutional network (FCN) and a transformer base with few-shot learning algorithms to extract information from spectral-spatial-temporal Sentinel-2 images. The attention mechanism in the transformer algorithm may effectively mitigate the issue of limited labeled samples and enhance the accuracy of learning correlations between samples, resulting in more successful classification. The experimental findings demonstrate that the LSST-Former model achieves an overall accuracy of 99.59% and an Intersection-over-Union (IoU) score of 98.84% for detecting mangrove loss, and the validation of universal applicability achieves an overall accuracy of more than 92% and a kappa accuracy of more than 89%. LSST-Former demonstrates superior performance compared to state-of-the-art deep-learning models such as random forest, Support Vector Machine, U-Net, LinkNet, Vision Transformer, SpectralFormer, MDPrePost-Net, and SST-Former, as evidenced by the experimental results and accuracy metrics.

**Keywords:** mangrove loss detection; few-shot learning; transformer; FCN



**Citation:** Panuntun, I.A.; Jamaluddin, I.; Chen, Y.-N.; Lai, S.-N.; Fan, K.-C. LinkNet-Spectral-Spatial-Temporal Transformer Based on Few-Shot Learning for Mangrove Loss Detection with Small Dataset. *Remote Sens.* **2024**, *16*, 1078. <https://doi.org/10.3390/rs16061078>

Academic Editor: Arturo Sanchez-Azofeifa

Received: 11 January 2024

Revised: 11 February 2024

Accepted: 12 March 2024

Published: 19 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Mangrove forests are crucial in supporting biodiversity, protecting coastlines, and mitigating climate change [1–4]. These unique ecosystems are home to a diverse variety of flora and fauna, with many species exhibiting specialized adaptations to survive in the challenging conditions of the intertidal zone [5]. Mangroves are vital habitats for many marine organisms, fish, crabs, and other aquatic creatures [6]. In addition, the complex root systems of mangroves play a crucial role in stabilizing coastal sediments, mitigating the effects of erosion and storm surges [7]. Mangrove forests have a substantial capacity for carbon sequestration, rendering them very helpful in reducing the effect of climate change [8].

Mangrove forests, although ecologically significant, face multiple threats from natural disasters and human activities. Extreme climate events [9], sea level rise [10], hurricanes [11], earthquakes [12], tornados [13], and tsunamis [14] are some of the natural disasters that have caused mangrove forest loss. Meanwhile, aquaculture, hydrological pollution, agriculture, timber extraction, urban development, and port development are some human-caused factors contributing to the loss of mangrove forests [15–17].

The accurate identification of mangroves is crucial for comprehending the current state of their development and decline, and it is highly valuable for mapping the mangrove ecosystem. Studies have shown that 35% of the world's mangroves disappeared between 1980 and 2000 [18], and an additional 2.1% perished between 2000 and 2016 [19]. Furthermore, the yearly rate of mangrove loss was estimated to be between 0.26% and 0.66% from 2000 to 2012 [20]. The primary obstacle in mangrove monitoring is addressing the degraded state of mangrove forests and their extensive dispersion within a given area.

The biggest challenge for mangrove mapping and monitoring is related to the field condition and the large area of mangrove ecosystems. The field condition of the mangrove ecosystem is muddy [21] and is affected by tides [22] and dense forests [23], as well as other conditions. These are the biggest challenges for direct mangrove surveys in large areas [24]. In recent years, satellite imagery has become an important tool for monitoring mangrove ecosystems worldwide. It offers continuous coverage over large areas, frequent revisits, and valuable spectral information. For mangrove loss mapping, the temporal resolution (frequent re-visits) of satellite imagery is the main advantage, because this can provide a satellite image that shows the mangrove condition in the past. Satellite imagery is helpful for various purposes related to mangroves, such as monitoring their habitats [25], distinguishing different species within them [26], and detecting changes in these areas over time [27]. The utilization of satellite imagery has facilitated the evaluation of the scope of mangroves, the mapping of their distributions, the analysis of factors causing changes, the identification of deterioration, and the provision of information for management methods.

In recent years, numerous researchers have utilized various methods to gain insights into the status of mangrove ecosystems. Chen et al. [22] used the Red-Edge Mangrove Index (REMI) with Sentinel-2 imagery for mapping mangroves on Hainan Island, China. To map and estimate mangrove areas in Qi'ao Island, China, Zhu et al. [23] used digital surface models and WorldView-2 Images. Hu et al. [28] mapped global mangrove AGB using multi-source satellite imagery, spaceborne LiDAR, and ground inventory data. Zhao et al. [29] mapped mangroves in China using Google Earth images, Sentinel-2 images, and Sentinel-1 data. Sharifi et al. [30] mapped mangrove forests in Qeshm Island, Iran using Sentinel-1 and Sentinel-2 satellite images.

Traditional approaches for obtaining information on the condition of mangrove ecosystems relied on the labor-intensive and time-consuming process of manually interpreting satellite imagery, resulting in high costs [31–33]. However, deep-learning models have demonstrated significant efficacy in automating this procedure, facilitating extensive surveillance and producing precise and rapid data [34,35]. Integrating satellite imagery with machine learning or deep learning might enhance efficiency and cost-effectiveness compared to traditional methods.

Recently, deep learning (DL) has frequently been used in the analysis of remote-sensing satellite imagery [36–39]. As an important part of machine learning characterized by utilizing multiple layers for processing, DL can improve model performance and accuracy [40]. Various deep-learning techniques have been developed to understand the condition of mangrove ecosystems. For example, Jamaluddin et al. [41] used MDPrePost-Net to determine the extent of mangrove degradation caused by the impact of Hurricane Irma as well as training data of around 40 million pixels with Sentinel-2 images and achieved an overall accuracy of 99.44%. Lin et al. [42] used Convex Deep Mangrove Mapping (CODE-MM) to map mangroves in several countries, along with Sentinel-2 images and training data of around 4–50 million pixels, to achieve an overall accuracy of 86.16–97.65%. Iovan et al. [43] used a Deep Convolutional Neural Network for mangrove mapping in Fiji, South Pacific

Ocean with Sentinel-2 and World-View 2 imagery. Diniz et al. [44] mapped mangroves using a random forest (RF) model, with Landsat-8, Landsat-7, and Landsat-5 imagery. Chen et al. [45] mapped mangroves in Dongzhaigang, China using a random forest (RF) model with Sentinel-2 imagery and achieved an overall accuracy of 93.23%. Xue et al. [46] used a two-stream translating long short-term memory network (TSTLN) with Sentinel-2 imagery for mangroves in Maowei Sea, Dongzhai Port, and Quanzhou Bay, China and achieved an overall accuracy of 90–97%. Al Dogom et al. [47] used random forest (RF) with Landsat-7 and Landsat-8 imagery for spatiotemporal monitoring and mapping of shoreline changes and mangroves in Umm al-Quwain, UAE. Though the previously mentioned method based on deep learning has demonstrated good results, it remains constrained in its ability to effectively utilize extensive training data. The problem with training deep learning with satellite imagery is due to a lack of sufficient labels [48]. Thus, creating a large, labeled image for training data is time-consuming.

Transformer is a deep-learning architecture initially created for natural language processing applications. However, it is now being utilized in several fields; examples of tasks include text detection [49], object detection [50], and image classification [51]. Transformers have been extensively utilized in several applications in remote-sensing fields, including hyperspectral image classification [52], change detection [53], and HSI super-resolution [54]. Transformers are excellent at understanding the spatial, spectral, and temporal connections within remote-sensing images [53]. Another deep-learning algorithm that significantly contributes to the remote-sensing field is a fully convolutional network (FCN) algorithm. This is a powerful tool for tasks like semantic segmentation, where the goal is to classify each pixel in an image into a particular class [55]. State-of-the-art FCN algorithms include U-Net [56], LinkNet [57], FPN [58], PSPNet [59].

The previous deep-learning models have relied on extensive training data. However, the transformer algorithm has shown remarkable performance with few-shot learning in hyperspectral images [38,53], and the FCN algorithm is frequently employed in remote sensing. Few-shot learning is a technique in which a model can accurately classify the categories in a new dataset using only a few labeled samples during the processing of the dataset [60,61]. Sentinel-2 is an accessible passive remote-sensing system with global coverage that has 13 bands with varying spatial resolutions (10, 20, and 60 m). Several studies have effectively utilized machine-learning and deep-learning methods with Sentinel-2 data to understand the condition of mangroves [22,24,27–30,39–42]. The FCN algorithm can extract more refined deep features and contribute to determining the spatial-spectral pattern of the images. Transformer algorithms are computational models that can effectively represent and process extended sets of information. The major contributions of this article for mangrove loss detection under limited labels can be highlighted as follows:

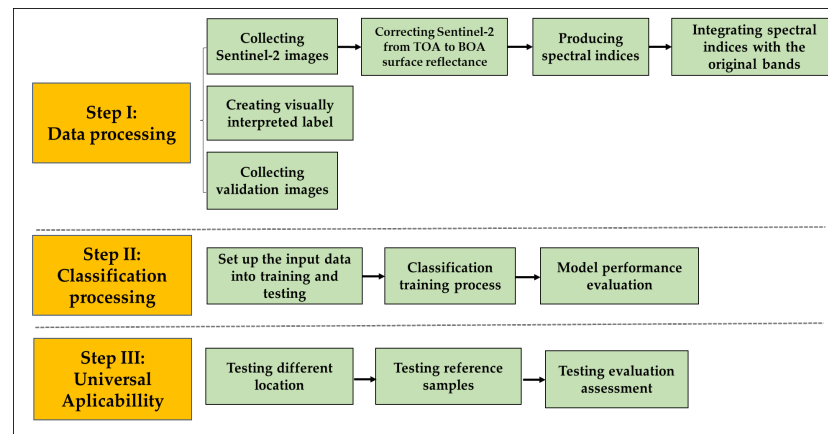
1. A LSST-Former method is proposed for an improved deep-learning model that only requires a few labeled samples by innovatively combining the FCN algorithm with a transformer and incorporating spatial, spectral, and temporal data from Sentinel-2 images to detect mangrove loss.
2. Experimental results strongly demonstrate the exceptional efficacy of our approach compared to other current models.
3. An analysis of the universal applicability of LSST-Transformer algorithms across different locations of the mangrove ecosystem is given.

The structure of this article is as follows: A discussion of the materials and methods is included in Section 2. The findings regarding mangrove loss detection under limited labeling are elaborated upon in Section 3. An analysis of our research is presented in Section 4. Finally, the author presents the final conclusions in Section 5.

## 2. Materials and Methods

The study's work stages are separated into three steps, as shown in Figure 1: (I) data processing, (II) classification processing, and (III) universal applicability. The data processing included collecting and correcting Sentinel-2 images from TOA to BOA surface

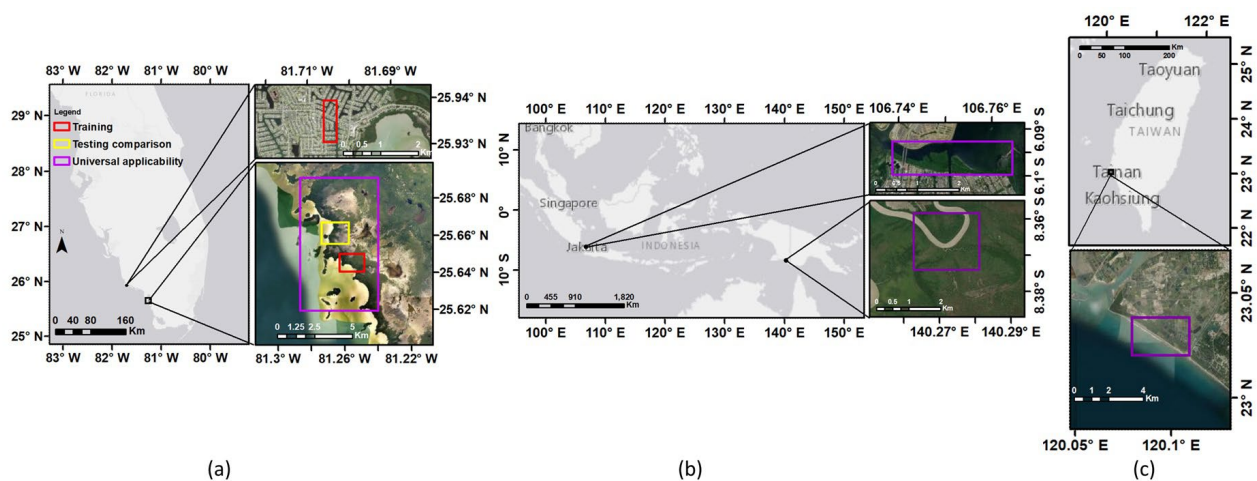
reflectance, producing spectral indices and integrating them with the original bands, creating visually interpreted labels, and collecting reference samples for testing evaluation assessments. The classification processing involves setting up the input data for training and testing, conducting the classification training procedure, and model performance evaluation. We proposed LSST-Former and trained it using input data (Section 2.4). The evaluation assessment was separated into two parts. The first part involved calculating the algorithm's output using the testing data from the classification process (Section 2.5). The second part involved calculating the different image locations from the classification process using the collected validation images (Section 2.6).



**Figure 1.** Scheme of all work stages in this research.

### 2.1. Study Area

Mangrove forests are primarily found inside the tropical and sub-tropical zones, specifically between the latitudes of  $30^{\circ}\text{N}$  and  $30^{\circ}\text{S}$ . The coastal zones chosen for analysis are located in southwest Florida, a mangrove area that can protect the land from natural disasters such as Hurricane Irma [41]; the Papua region of Indonesia, which has roughly 13.4% of the global carbon stores in mangroves [62]; PIK, Jakarta, Indonesia, a region that protects the land from seawater flooding, seawater intrusion, and abrasion from the north coast of Jakarta [63]; and Tainan, Taiwan, a region with a rich variety of wildlife and plant life, including a rare migratory bird species known as the black-faced spoonbill [64], as shown in Figure 2. The coastal zone of southwest Florida is the key investigation area, where Hurricane Irma, a Category 3 storm, passed through in September 2017 and caused significant loss to the local mangrove environment.



**Figure 2.** Locations of the study areas: (a) southwest Florida, (b) PIK, Jakarta and Papua, Indonesia, and (c) Tainan, Taiwan.

## 2.2. Satellite Data and Preprocessing

The satellite data utilized in this study include Sentinel-2 images provided by the European Space Agency (ESA), which were acquired between 2016 and 2019 and downloaded from the Google Earth Engine [65]. The properties of Sentinel-2 images are presented in Table 1. A total of 14 images were used in this study: seven images captured before mangrove ecosystem loss and seven images captured after mangrove ecosystem loss. A summary of the longitude and latitude coordinates, retrieval year, and picture size of the Sentinel-2 data used in the experimental investigation is presented in Table 2. The date of the image data is chosen based on the criterion of minimal cloud cover, particularly at the specific study location. The southwest Florida dataset was chosen for the training model. In September, southwest Florida encountered Hurricane Irma, which was highly valuable for developing a classification model for mangrove loss detection in this period. In addition, the purpose is to observe the spatial variations in mangroves in different regions.

**Table 1.** Properties of Sentinel-2 images.

Band	Band Name	Central Wavelength (nm)	Spatial Resolution
B1	Aerosol	442.3	60
B2	Blue	492.1	10
B3	Green	559	10
B4	Red	665	10
B5	Red Edge 1	703.8	20
B6	Red Edge 2	739.1	20
B7	Red Edge 3	779.7	20
B8	Near-infrared (NIR)	833	10
B8A	Red Edge 4	864	20
B9	Water-vapor	943.2	60
B10	Cirrus	1376.9	60
B11	Short-wave infrared (SWIR 1)	1610.4	20
B12	Short-wave infrared (SWIR 2)	2185.7	20

**Table 2.** Latitude, longitude, date acquisition, image size, and usage of the Sentinel-2 data in our experiments.

Name (Scene Code)	(Lon) <sup>o</sup>	(Lat) <sup>o</sup>	Date Before	Date After	Image Size	Usage
Southwest Florida-1 (T17RMJ)	−81.263197 −81.2488207	25.6407337 25.6497094	20161001	20180104	145 × 100	Training/testing
Southwest Florida-2 (T17RMJ)	−81.7059548 −81.7026501	25.9303954 25.9393493	20161001	20180104	35 × 100	Training/testing
Southwest Florida-3 (T17RMJ)	−81.2741405 −81.2583906	25.6556281 25.6668073	20161001	20180104	159 × 125	Model applicability
Southwest Florida-4 (T17RMJ)	−81.286543 −81.2405811	25.6903579 25.6195264	20161001	20180104	464 × 786	Universal applicability
Southwest Florida-5 (T17RMJ)	−81.4133421 −81.413342	25.3594457 25.8375025	20161001	20180104	3000 × 5276	Universal applicability
PIK Jakarta (T48MXU)	106.7384884 106.7644093	−6.1047809 −6.0976117	20190505	20191101	289 × 81	Universal applicability
Papua (T54LVR)	140.2627755 140.2871514	−8.3748976 −8.3514604	20171104	20181005	270 × 260	Universal applicability
Tainan (T50QRL)	120.079663 120.1093595	23.0201424 23.0383102	20180302	20180903	310 × 209	Universal applicability

The Sentinel-2 satellite contains a total of 13 bands, which have three different spatial resolutions. There are 4 bands with a spatial resolution of 10 m, 6 bands with a spatial resolution of 20 m, and 3 bands with a spatial resolution of 60 m. We selected Level-1C imagery, which refers to Top-of-Atmosphere (TOA) products that have gone through radiometric and geometric correction to align with a global reference system. Using the SIAC atmospheric correction module [66], we converted Level-1C data into Level-2A images, specifically orthoimages that have been corrected for Bottom-of-Atmosphere (BOA) reflectance. Upon receiving the Sentinel-2 Level-2A products, we adjusted the spatial resolution of the SWIR-1 and SWIR-2 data to a 10 m spatial resolution, because calculating mangrove indices and the deep-learning model input require consistent spatial resolution.

The input bands and spectral indices for the training model are based on previous research on mangrove mapping [41,44–46]. Spectral indices are mathematical formulas that are frequently used to evaluate the bands of remote-sensing images in order to enhance the ability to distinguish various objects of interest according to their spectral properties. Four spectral indices were computed: the modular mangrove recognition index (MMRI) [44], the normalized difference mangrove index (NDMI) [67], the combined mangrove recognition index (CMRI) [68], and the normalized difference vegetation index (NDVI) [69]. The formulas are demonstrated in Table 3.

**Table 3.** Formulas for spectral indices.

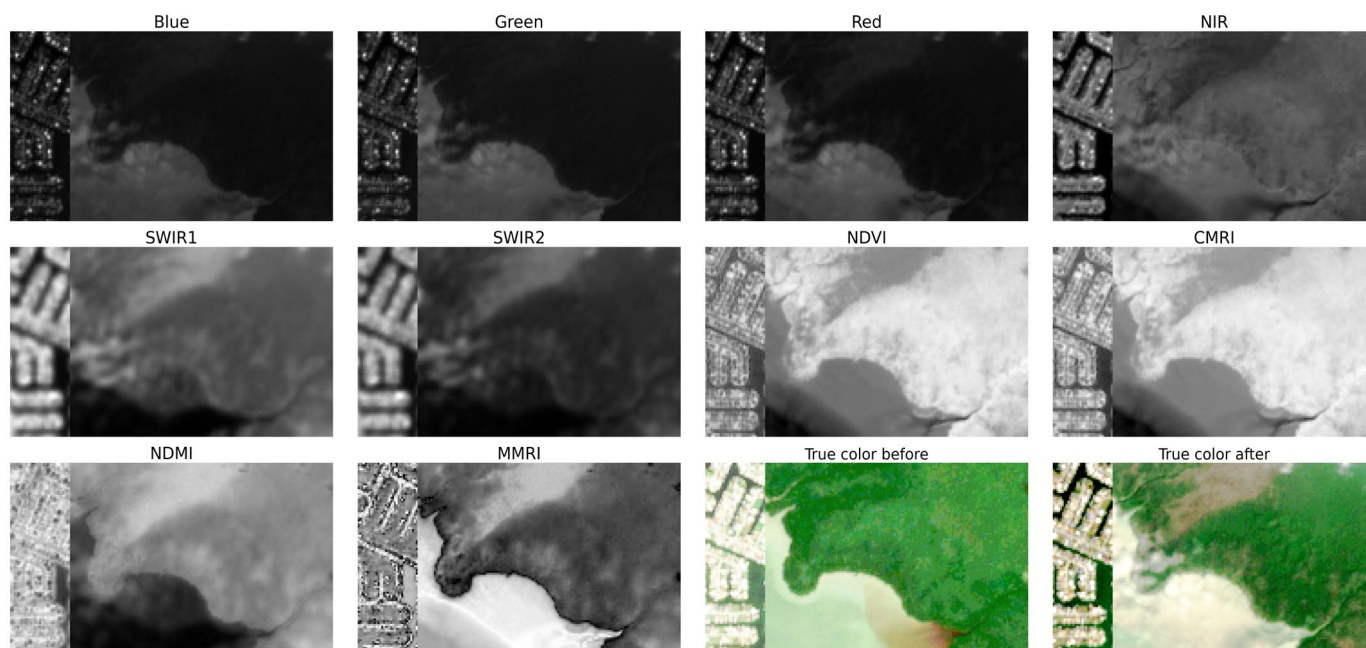
Total Label (Pixel)	Formula	Reference
NDVI	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$	[69]
NDWI	$(\text{Green} - \text{NIR}) / (\text{Green} + \text{NIR})$	[70]
CMRI	$(\text{NDVI} - \text{NDWI})$	[68]
NDMI	$(\text{SWIR2} - \text{Green}) / (\text{SWIR2} + \text{Green})$	[67]
MNDWI	$(\text{Green} - \text{SWIR1}) / (\text{Green} + \text{SWIR2})$	[71]
MMRI	$( \text{MNDWI}  -  \text{NDVI} ) / ( \text{MNDWI}  +  \text{NDVI} )$	[44]

These indices can contribute to defining objects as either mangrove or non-mangrove based on previous findings. The normalized difference vegetation index (NDVI) is frequently employed in remote sensing to assess varied vegetation objects, including mangrove objects [41,46]. The CMRI is obtained by analyzing the difference between the normalized difference water index (NDWI) values [70] and the normalized difference vegetation index (NDVI) values. This process helps to distinguish between mangroves and non-mangrove objects; it can address seasonal features and improve the pattern of mangrove objects within the output result. The NDMI was created to improve upon the disparity between mangrove and non-mangrove vegetation types in satellite imagery by leveraging the normalization difference between the green and SWIR-2 bands, which are connected with the pattern of mangrove objects. The MMRI was developed to improve the disparity in brightness between mangrove and non-mangrove objects, allowing for recognition of mangrove objects. The NDVI and modified NDWI (MNDWI) [71] are combined to create the MMRI. This study utilized a total of 10 input bands, which included the original band (blue, green, red, NIR, SWIR-1, SWIR-2) and four spectral indices (NDVI, CMRI, NDMI, and MMRI), as shown in Figure 3.

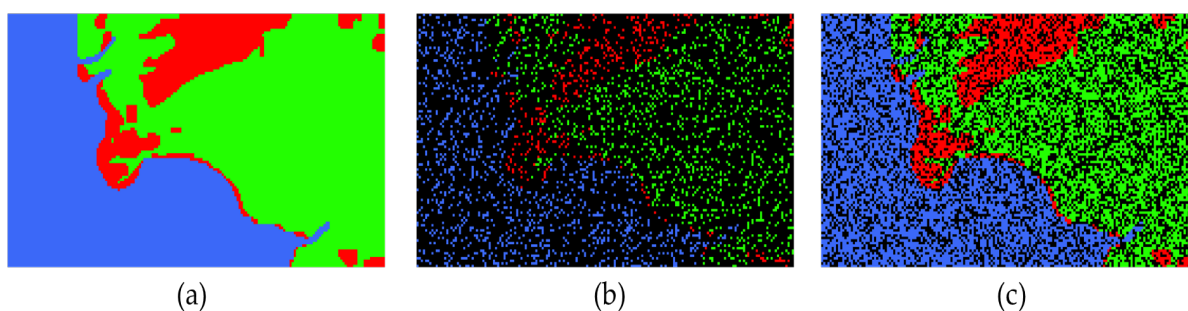
### 2.3. Input Data for Model

The training data utilized for the model consisted of pixel-level samples, where each pixel in the sample was assigned a specific attribute value. The original remote-sensing imagery from Sentinel-2 was manually annotated with high-resolution photos obtained from Google Earth, which were used to produce visually interpreted labels for the areas before and after the loss occurred. The labels were visually interpreted and divided into three categories: (1) intact mangroves, which refers to mangrove objects that remained unchanged in the images taken before and after; (2) mangrove loss, which refers

to mangrove objects that were intact in the initial images but disappeared or degraded in the subsequent images; and (3) non-mangrove areas, which includes all objects that were not classified as mangrove objects, such as water bodies, buildings, bare land, and non-mangrove vegetation. Prior studies [39,40] have employed large training data of 4–50 million pixels, whereas our research utilizes small training data that are typically a hundred to ten thousand times smaller [72]. We utilized the random shot method, as depicted in Figure 4, to select 2870 pixels for training data and 10,717 pixels for testing data. This approach was employed to obtain improved outcomes despite the constraint of few labeled data.



**Figure 3.** Ten input bands and true-color images.



**Figure 4.** (a) Visually interpreted labels in Southwest Florida-1 and Southwest Florida-2, (b) training data, and (c) testing data. Blue, green, and red are non-mangrove, mangrove, and mangrove loss, respectively.

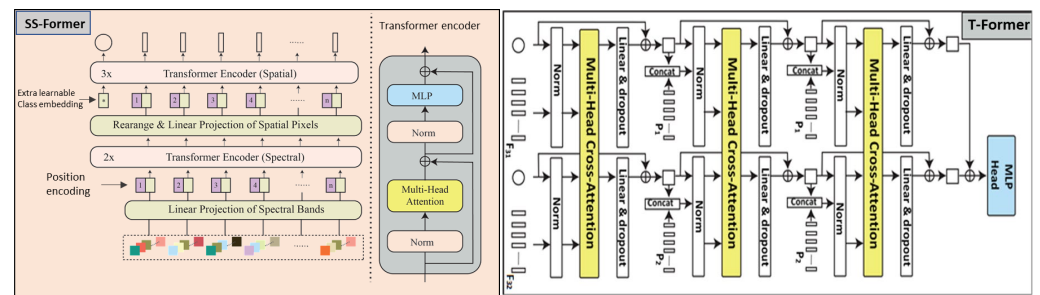
Table 4 shows the total number of training and testing pixels for each class in our study. The non-mangrove class has a greater overall pixel count, because it covers various land cover types (such as water bodies, non-mangrove vegetation, urban areas, open areas, etc.) and is a background class.

**Table 4.** Training and testing size.

Total Label (Pixels)	Training (Pixels)	Testing (Pixels)
Non-mangrove	1194	4549
Mangrove	1268	4608
Mangrove loss	408	1560
Total	2870	10,717

#### 2.4. LSST-Former Architecture

Generally, the proposed LSST-Former is a combination model of a FCN algorithm and a transformer algorithm; we adopted LinkNet architecture and SST-Former architecture, as shown in Figure 5. Firstly, we employ the FCN algorithm to train the data and obtain the FCN pre-trained model. Afterward, the input image is forwarded to the pre-trained fully convolutional network (FCN) model to generate a feature map. This feature map is then concatenated with the original image. Finally, the image concatenation is randomly sampled and sent to the SST-Former to obtain the final feature map. The details about sub-models in the proposed LSST-Former are introduced in Sections 2.4.1 and 2.4.2.

**Figure 5.** Architecture of LSST-Former. SS-Former and T-Former are adopted from [48].

##### 2.4.1. FCN Feature Extractor

The FCN feature extractor in LSST-Former is a spatial-spectral extractor specifically designed to assess the status of mangroves. This extractor utilizes the LinkNet [57] architecture as the convolutional network to enhance the differentiation between various objects of interest, relying on their spatial-spectral features. The sub-model comprises encoder and decoder components.

We utilized the VGG-16 design to build the encoder component by eliminating two complete connection layers and the SoftMax function from the VGG-16 architecture. The initial and subsequent encoder blocks comprise two 2D convolutional layers and the activation function utilizing the Rectified Linear Unit (ReLU), followed by a 2D max-pooling layer. The remaining three encoder blocks comprise three 2D convolutional layers and the activation function utilizing ReLU, followed by a 2D max-pooling layer. The lower section of this sub-model, located after the encoder section, has a feature map size of  $2 \times 2$ . It comprises two 2D convolutional layers, which are followed by batch normalization and ReLU activation. The decoder component employs the Upsampling2D layer to increase the resolution of the feature map. The decoder component comprises two 2D convolutional layers, Upsampling2D, ReLU, and batch normalization.

This sub-model utilized a skip connection with an add operation layer between each block of the encoder and decoder. This is demonstrated in Equation (1).

$$X_{sc,i} = X_{ec,i} + X_{dc,i}, i \in \{1, 2, \dots, 5\} \quad (1)$$

where  $X_{sc}$  represents the skip connection,  $X_{ec}$  represents the encoder block, and  $X_{dc}$  represents the decoder block. The last 2D convolutional layer employs a kernel size of  $1 \times 1$  and utilizes the SoftMax activation algorithm.



$$X_{out,i} = softmax(X_{sc,5}), i \in \{1, 2, \dots, n\} \tag{2}$$

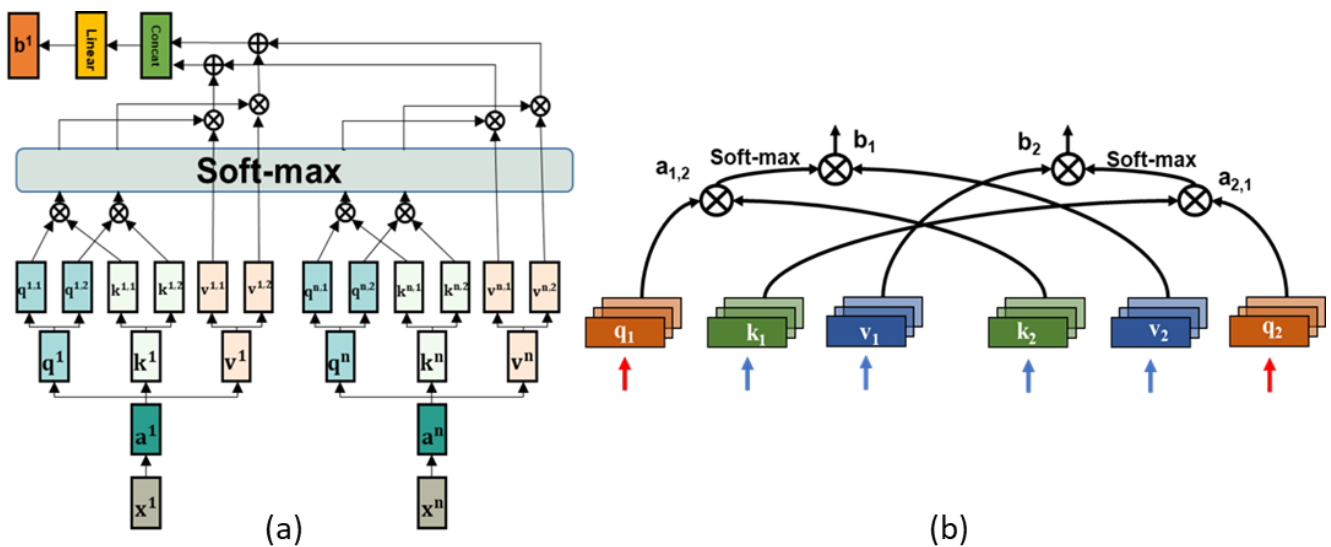
where  $X_{out,i}$  is the output of the FCN feature extractor.

### 2.4.2. Transformer Classifier

The LSST-Former includes a transformer architecture as the last classifier in its second sub-model. We adopted SST-Former architecture [53]. This sub-model consists of SS-Former and T-Former components, and we modified it with a soft cross-entropy loss function [73]. The transformer structures have three primary components: position embedding, attention, and multi-layer perceptron (MLP). Position embedding helps reduce network complexity. The attention mechanism module learns and categorizes relationships by comparing and scoring similarities between different samples. High scores indicate that they are in the same class; low scores indicate that they are in different classes. Meanwhile, the MLP is employed as a feedforward network within the transformer encoder, and it is classified in the transformer decoder.

The sequential order is encoded as an input series using position embedding. Vaswani et al. [74] employed position embedding using sine and cosine functions with different frequencies. Gehring et al. utilized position embedding using trainable weights [75]. The input items  $x = (x_1, \dots, x_n)$  were embedded in the distribution space  $w = (w_1, \dots, w_n)$ , where  $n$  represents the number of input components. The model was then given positional embedding  $p = (p_1, \dots, p_n)$ , and the two were merged to yield the input element representation  $e = (w_1 + p_1, \dots, w_n + p_n)$ . Model training was used to determine  $e$  and  $p$ .

Attention is a crucial component of the transformer model, primarily employed to ascertain the similarity between various samples and facilitate classification. The objective of attention is to acquire the similarity among any two training samples. The relationship score between different samples is determined by assessing their similarities. Relationship scores are higher for samples within the same category and lower for samples between different categories. By employing this approach, the attention mechanism may effectively capture the relational information among samples, leading to enhanced classification accuracy and model robustness. The SST-Former model consists of two attention mechanisms, as shown in Figure 6: multi-head self-attention (MSA) in the SS-Former part and multi-head cross-attention (MCA) in the T-Former part.



**Figure 6.** (a) Multi-head self-attention (MSA), (b) multi-head cross-attention (MCA) modified from [53],  $\otimes$  indicates dot product. This is a simple example of  $n$  inputs and one output, and the output of other elements involves the same process.  $q$ ,  $k$ , and  $v$  refer to query, key, and value, respectively.

The transformer encoder utilizes a feedforward network employing a MLP. This MLP comprises linked layers and employs a rectified ReLU activation function. The MLP also serves a non-linear function. Another function within the transformer decoder is the classification task, which is referred to as the MLP head. The MLP head consists of a fully linked layer and a layer normalization. Hence, MLP is essential for transformers.

The training and testing process involves patch pairs, where  $X = \{X_1, X_2\}$ . For one branch, we reshape a patch  $X \in R^{w \times h \times c}$  into a spectral sequence set  $x = \{x_1, x_2, \dots, x_c\}$ , where  $w, h$  represent the spatial resolution of the patch, and  $c$  represents the number of channels. In order to simplify the calculation process, we regard  $x$  as a matrix of  $c \times wh$ .

The SST-Former method for mangrove loss detection is a sophisticated process that involves several key steps.

The process begins with linear projection and position encoding. In this step, the SST-Former position encodes each pixel on the cube. This is a crucial step, as it allows the model to remember the spectral and spatial sequences of each pixel.

$$E = x\omega_1, \omega_1 \in R^{wh \times n} \quad (3)$$

$$Z = E + p, E \in R^{c \times n}, p \in R^{c \times n} \quad (4)$$

where  $E$  represents linear projection,  $\omega_1$  represents the weight matrix,  $p$  represents position embedding, and  $Z$  represents output position embedding. Next, the spectral transformer encoder comes into play. This structure is specifically designed to extract spectral sequence information. It processes the spectral data of the spectral images.

$$H = TE(Z), H \in R^{c \times n} \quad (5)$$

where  $TE$  represents the transformer encoder, and  $H$  represents the output spectral transformer encoder. Then, we reshape  $H \in R^{c \times n}$  into  $H \in R^{n \times c}$ . Following this, a class token  $C \in R^{1 \times m}$  is used. This token serves to store the class information of a single temporal image. Interestingly, this token concatenates the output of the spectral transformer encoder, effectively combining the spectral information with the class information.

$$E = H\omega_2, \omega_2 \in R^{c \times m} \quad (6)$$

$$Z = E + p, E \in R^{n \times m}, p \in R^{n \times m} \quad (7)$$

$$G = \text{Concat}(C, Z) \quad (8)$$

The spatial transformer encoder is then used. This encoder is tasked with extracting spatial texture information. It focuses on the spatial data of the images, thereby complementing the spectral data processed by the spectral transformer encoder.

$$F_3 = TE(G), F_3 \in R^{(n+1) \times m} \quad (9)$$

Thus, a pseudo-category of patch can be observed. We can obtain two results,  $F_{31} = \{C_1, P_1\}$  and  $F_{32} = \{C_2, P_2\}$ , where  $F_3$  represents the output spatial transformer encoder,  $C$  represents the class token, and  $P$  represents the patch token.

Finally, the temporal transformer and MLP are utilized. The features of different temporal images are sent as the input of the temporal transformer. This transformer is used to extract useful mangrove loss detection features between the current image pairs. The result is then obtained through a multilayer perceptron (MLP).

$$y' = LN(TF\{[F_{31}, F_{32}]\})\omega, \omega \in R^{m \times 3}, y' \in R^{1 \times 3} \quad (10)$$

where  $y'$  is the final result. We modified the loss function with the soft cross-entropy loss function [73], as calculated by

$$CE = -\sum_{i=1}^C y(i)\log(y'(i)), \quad (11)$$

$$SCE = -\sum_{i=1}^C ((1 - \sigma)\delta_{i,y} + \frac{1}{C})\log(y'(i)), \quad (12)$$

where  $CE$  is cross-entropy,  $SCE$  is soft cross-entropy,  $y$  is the target,  $y'$  is the prediction,  $\sigma$  is the smoothing factor,  $C$  is the class, and  $\delta_{i,y}$  is the Dirac delta, which equals 1 for  $i = y$  and 0 otherwise.

### 2.5. Evaluation Assessment

The Intersection over Union (IoU) [76], overall accuracy (OA), and F1-Score [77] were utilized as quantitative indicators. They are calculated as follows:

$$IoU = \frac{|target \cap prediction|}{|target \cup prediction|} \quad (13)$$

$$OA = \frac{\sum_{i=1}^C TP_i}{\sum_{i=1}^C (TP_i + FP_i + TN_i + FN_i)} \quad (14)$$

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (15)$$

$$precision = \frac{TP}{TP + FP} \quad (16)$$

$$recall = \frac{TP}{TP + FN} \quad (17)$$

IoU is calculated as the area of overlap between the target and prediction results divided by the area of union between the target and prediction results. A true positive (TP) refers to the situation where the model accurately predicts the positive class. A true negative (TN) refers to the situation where the model accurately predicts the negative class. A false positive (FP) refers to the situation where the model makes an incorrect prediction by classifying an instance as positive when it is actually negative. On the other hand, a false negative (FN) refers to the situation where the model makes an incorrect prediction by classifying an instance as negative when it is actually positive.

### 2.6. Validation of Universal Applicability Model

The final task was to evaluate our model's efficacy in detecting mangrove loss under limited labels. We tested the model in several areas different from the training process. We utilized the model to analyze Southwest Florida-4, PIK Jakarta, Papua, and Tainan images to obtain the detection outcome. We employed stratified random sampling within the study zone to obtain prediction point samples for the precise calculation of the accuracy of the mangrove loss map. Prediction point samples were compared using historical ESRI World Imagery Wayback [78]. A total of 1500 prediction points were validated across Southwest Florida-4, 300 prediction points were validated across PIK Jakarta, Indonesia, 900 prediction points were validated across Papua, Indonesia, and 600 prediction points were validated across Tainan, Taiwan. Specifically, there were 500, 100, 300, and 200 prediction points for each class, respectively. A kappa coefficient (Kappa) was utilized for quantitative indicators, which are calculated as follows:

$$Kappa = \frac{OA - Pc}{1 - Pc}, \quad (18)$$

$$P_c = \frac{(TP + FP)(TP + FN)(TN + FN)(TN + FP)}{(TP + FP + TN + FN)^2}, \quad (19)$$

A true positive (TP) refers to the situation where the model accurately predicts the positive class. A true negative (TN) refers to the situation where the model accurately predicts the negative class. A false positive (FP) refers to the situation where the model makes an incorrect prediction by classifying an instance as positive when it is actually negative. On the other hand, a false negative (FN) refers to the situation where the model makes an incorrect prediction by classifying an instance as negative when it is actually positive.

### 2.7. Implementation Detail

The Sentinel-2 images were acquired before and after mangrove loss, consisting of 10 bands (blue, green, red, NIR, SWIR1, SWIR2, NDVI, CMRI, NDMI, and MMRI). These bands, together with visually interpreted labels, were initially selected as random shot samples and used as input data for our deep-learning model. We employed Adam as the optimizer, utilizing the default parameter configuration and a learning rate of 0.001, as suggested in the original Adam paper. The training process consisted of 400 iterations, utilizing 2870 input data for training and 10,717 for testing. The training loss during training was derived using the soft cross-entropy loss. This approach employs the Python programming language along with the Tensorflow, Keras, and PyTorch frameworks. The Python code was run on a Windows 11 operating system using an Intel Core i9-11900K central processing unit, an NVIDIA GeForce RTX 3090 graphics processing unit, and 64 GB RAM.

## 3. Results

The following section showcases the experimental findings of the LSST-Former (Section 3.1). The LSST-Former results are compared with a well-established architecture in Section 3.2. We examined the input data when integrated with vegetation and mangrove indices including CMRI, NDVI, MMRI, and NDMI. We discovered that including these spectral indices enhanced the accuracy of the results (Section 3.3). We evaluate the effects of the parameters of LSST-Former in Section 3.4. Finally, we applied the proposed LSST-Former for universal applicability in Section 3.5.

### 3.1. LSST-Former

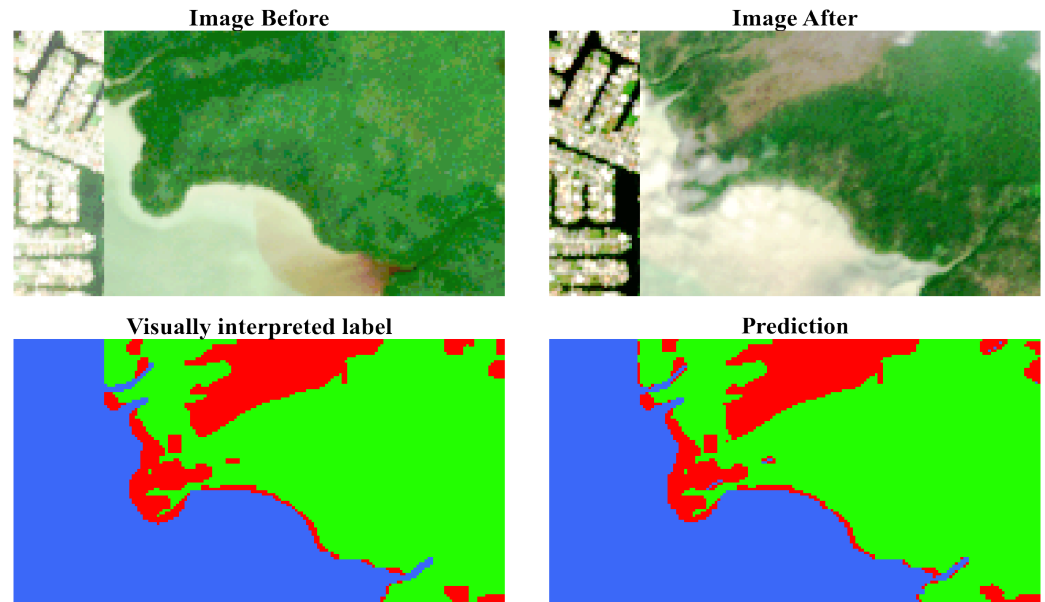
The LSST-Former's trained model was assessed using a distinct testing dataset, which differed from the training data used for input. The proposed method for learning with low data utilizes only 2870 training samples for LSST-Former. Given that the effectiveness of deep-learning methods depends greatly on the quantity of data at hand, a direct demonstration of the advantages of our small-data learning strategy is the remarkable performance attained by LSST-Former utilizing only 2870 restricted training samples. The performance metrics for areas that are not mangroves, areas with intact mangroves, and areas with mangrove loss can be seen in Table 5. A visual representation of these results can be observed in Figure 7.

**Table 5.** Quantitative classification results from 10,717 pixels of testing data.

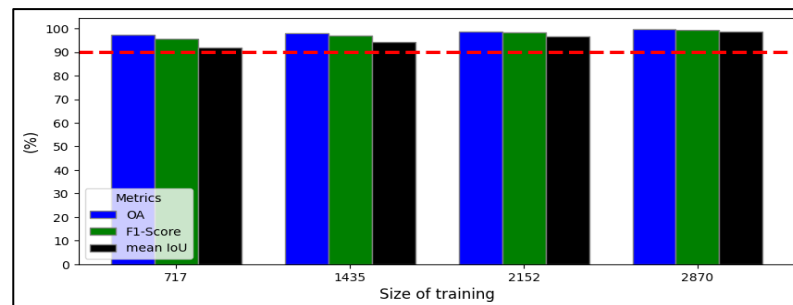
Metrics	Non-Mangrove	Intact Mangrove	Mangrove Loss
IoU	99.62	99.33	97.59
F1-Score	99.81	99.66	98.78
Precision	99.91	99.58	98.72
Recall	99.71	99.74	98.84

The impacts of different total numbers of labels used for training are illustrated in Figure 8 and tabulated in Table 6. To investigate the impact of the total number of input data on the accuracy and IoU score, we conducted experiments utilizing label sizes of

717, 1435, 2152, and 2870 pixels for training data. The experimental results demonstrate a gradual increase in both the overall accuracy and IoU scores as the total number of input data increases and the dashed line show evaluation assessment more than 90%. For this investigation, we ultimately used 2870 pixels as the definitive quantity of labeled data for training.



**Figure 7.** Visual results of entire training and testing region in Southwest Florida-1 and Southwest Florida-2. Blue, green, and red are non-mangrove, mangrove, and mangrove loss, respectively.



**Figure 8.** The impact of various training sizes on 10,717 pixels of testing data.

**Table 6.** The impact of various training sizes on 10,717 pixels of testing data.

Training Size	Overall Accuracy	F1-Score	Mean IoU
717	97.18	95.74	92.05
1435	98.02	97.02	94.34
2152	98.81	98.21	96.54
2870	99.59	99.41	98.84

Ablation tests were undertaken to assess the efficacy of each component in LSST-Former. In order to assess the impact of each component on the overall architecture, we selectively eliminated different elements of LSST-Former and analyzed the resulting accuracy. The classification outcomes of eliminating components in the proposed model are presented in Table 7. The findings indicate that excluding LinkNet from our model resulted in a 5.74% decrease in mean IoU, a 3.07% decrease in F1-score, and a 2.01% decrease in overall accuracy. Conversely, removing SST-Former from our model led to

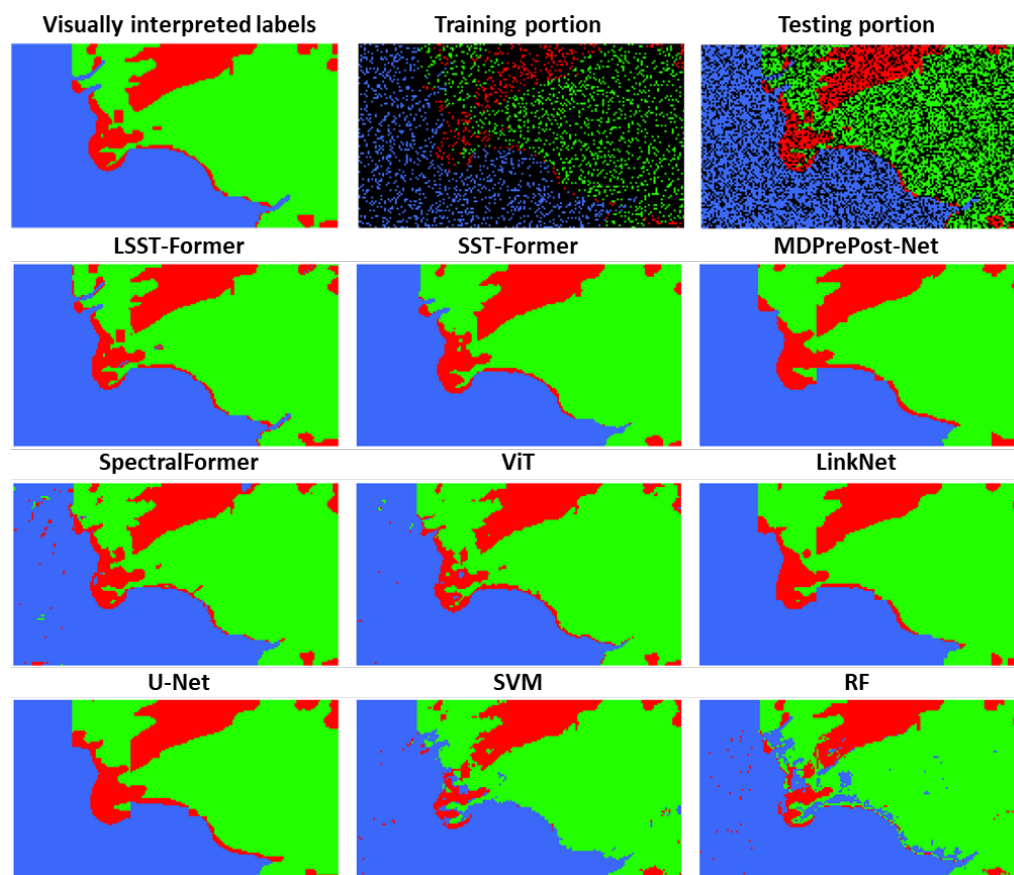
a 14.29% decrease in mean IoU, an 8.38% decrease in F1-score, and a 3.68% decrease in overall accuracy.

**Table 7.** Analysis of each removed part in LSST-Former.

Architecture	Overall Accuracy	F1-Score	Mean IoU
LSST-Former	99.59	99.41	98.84
No LinkNet	97.58	96.34	93.10
No SST-Former	95.81	91.03	84.55

### 3.2. Comparison with Other Well-Established Architectures

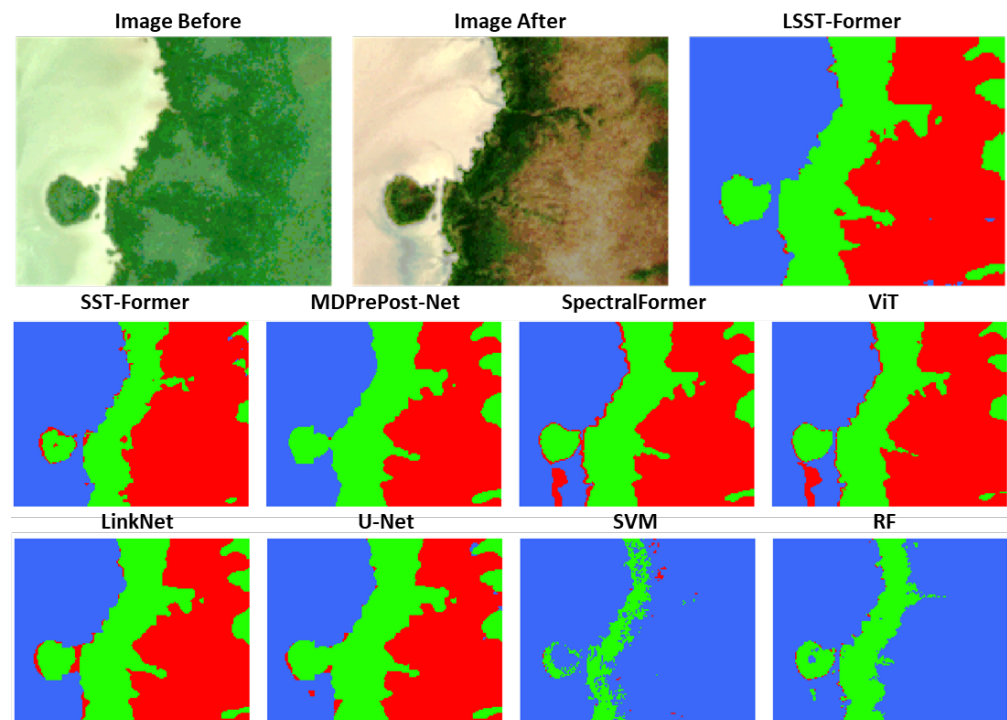
The performance of the suggested model was evaluated by comparing it to various established architectures, such as random forest (RF) and Support Vector Machine (SVM), U-Net [56], LinkNet [57], Vision Transformer (ViT) [51], SpectralFormer [38], MDPrePost-Net [41], and SST-Former [53]. The RF, SVM, ViT, SpectralFormer, and SST-Former models were trained using the same input data of 2870 samples. In contrast, the U-Net, LinkNet, and MDPrePost-Net models were trained using patch images of size  $32 \times 32$ , generated using the patchify Python module with a step size of 16 pixels. Figure 9 visually compares the LSST-Former output and several established architectural models. According to the visualization analysis, the proposed model obtained a superior classification performance compared to other existing architectures.



**Figure 9.** Visual comparison of LSST-Former with other architectures in Southwest Florida-1 and Southwest Florida-2. Blue, green, and red are non-mangrove, mangrove, and mangrove loss, respectively.

Traditional classification techniques such as random forest (RF) and Support Vector Machine (SVM) produce salt-and-pepper noise in classification maps, implying that these classifiers fail to identify the materials of objects accurately. Convolutional neural net-

work models, such as U-Net and LinkNet, provide smooth classification maps using their robust nonlinear data-fitting capabilities. However, these models still have significant underestimation and overestimation areas. Transformers like ViT and Spectralformer are a developing network architecture that can effectively mitigate huge underestimations and overestimations, even though they still have salt-and-pepper noise. Utilizing the temporal image convolution-based models MDPrePost-Net and SST-Former can effectively minimize misclassification. Our proposed LSST-Former model obtains highly desirable classification maps and less misclassification than other architectures. Figure 10 show a comparison of model applicability with a Southwest Florida-3 image.



**Figure 10.** Model applicability with Southwest Florida-3 image. Blue, green, and red are non-mangrove, mangrove, and mangrove loss, respectively.

The classification results are presented in Table 8, including three main metrics: mean IoU, F1-Score, and overall accuracy. Additionally, the IoU for each individual class is also reported. The LSST-Former architecture surpasses other architectures in terms of the mean Intersection-over-Union (IoU) and IoU scores, the F1-Score, and the overall accuracy for each individual class.

**Table 8.** Quantitative comparison of LSST-Former with other architectures in testing portion. Mg, Non-Mg, and MgLs are intact mangrove, non-mangrove, and mangrove loss class, respectively.

Class	Non-Temporal Imagery					Temporal Imagery			
	Conventional Network		Convolution Network		Transformer Network	Convolution Network		Transformer Network	
	RF	SVM	U-Net	LinkNet	Vit	Spectralformer	MDPrePost-Net	SST-Former	LSST-Former
IoU Non-Mg	81.33	88.33	92.41	92.42	96.04	95.71	93.06	98.39	<b>99.62</b>
IoU MgLs	55.60	57.40	54.80	65.79	76.69	82.29	67.44	85.33	<b>97.59</b>
IoU Mg	90.27	94.37	93.57	95.44	94.58	96.32	94.56	95.58	<b>99.33</b>
Mean IoU	75.74	80.04	80.26	84.55	89.10	91.44	85.02	93.10	<b>98.84</b>
F1-Score	85.35	87.94	87.84	91.03	86.81	95.41	91.39	96.34	<b>99.41</b>
OA	90.84	93.64	94.42	95.81	96.20	96.96	95.71	97.58	<b>99.59</b>

### 3.3. The Impact of Mangrove and Vegetation Indices

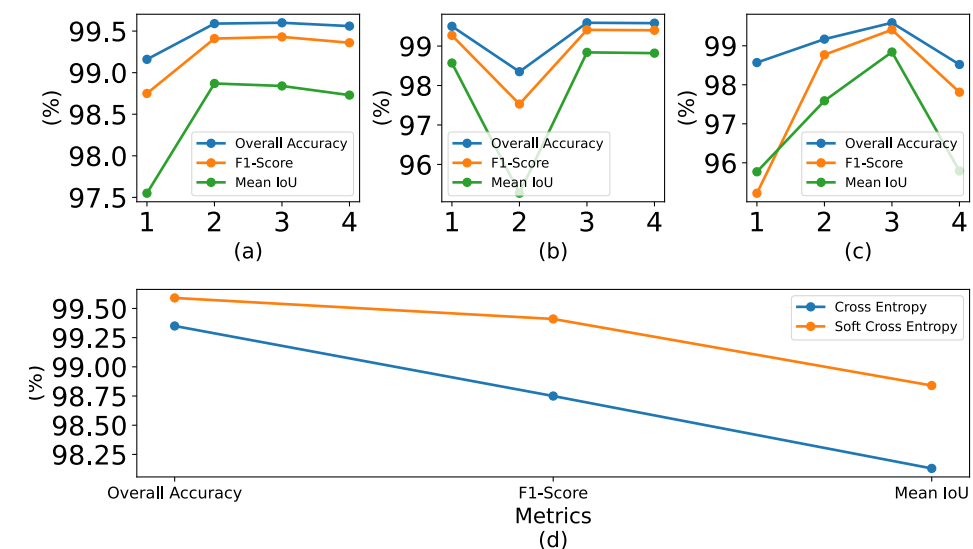
This section illustrates the impact that modifying the overall quantity of input bands has on LSST-Former. During the preliminary evaluation, the input data consisted solely of RGB bands obtained from Sentinel-2. RGB images are frequently employed in classifying images using deep-learning techniques. During the second experiment, RGB images were utilized that incorporated the near-infrared band consisting of wavelengths ranging from blue to red. During the third experiment, RGB images were employed to incorporate the near-infrared and short-wave infrared bands. Specifically, the blue, green, red, near-infrared, short-wave infrared 1, and short-wave infrared 2 bands were utilized. In the third experiment, the impact of employing short-wave infrared (SWIR) bands on the distinction between undisturbed mangrove areas and those that have experienced mangrove loss was evaluated. This is due to the fact that the SWIR spectrum is particularly sensitive to moist objects. The last experiment demonstrated that incorporating the 10 input bands, including the MMRI, NDMI, CMRI, and NDVI, enhances the effectiveness of classifying mangroves and detecting mangrove decline using the initial input data. The accuracy evaluation outcomes, comprising the average overall accuracy (OA), F1-Score, and mean Intersection over Union (IoU), are displayed in Table 9. After examining the input data using vegetation, wet objects, and mangrove indices, it is clear that using these indices greatly improved the accuracy of the classification.

**Table 9.** Accuracy results of added vegetation and mangrove indices.

Architecture	Overall Accuracy	F1-Score	Mean IoU
RGB	94.31	92.01	85.72
RGB NIR	94.70	91.10	86.71
RGB NIR SWIR1 SWIR2	95.03	93.13	87.53
<b>All</b>	<b>99.59</b>	<b>99.41</b>	<b>98.84</b>

### 3.4. Effects of Parameters

The efficacy of LSST-Former is based on the quality of the parameters. Hence, the impact of parameters such as the number of spectral transformer encoders, the number of spatial transformer encoders, the number of temporal CA layers, and the loss function are examined. The mean Intersection-over-Union (IoU), F1-Score, and overall accuracy (OA) findings obtained in our investigation are presented in Figure 11.



**Figure 11.** Accuracy of SST-Former with different parameters: (a) number of spectral transformer encoders; (b) number of spatial transformer encoders; (c) number of temporal CA layers; and (d) comparison of loss function.



**Optimal spectral transformer encoders:** Four encoder quantities were evaluated in order to determine the optimal number for a spectral transformer encoder set: one, two, three, and four. As illustrated in Figure 11a, the optimal number of spectral transformer encoder layers is two.

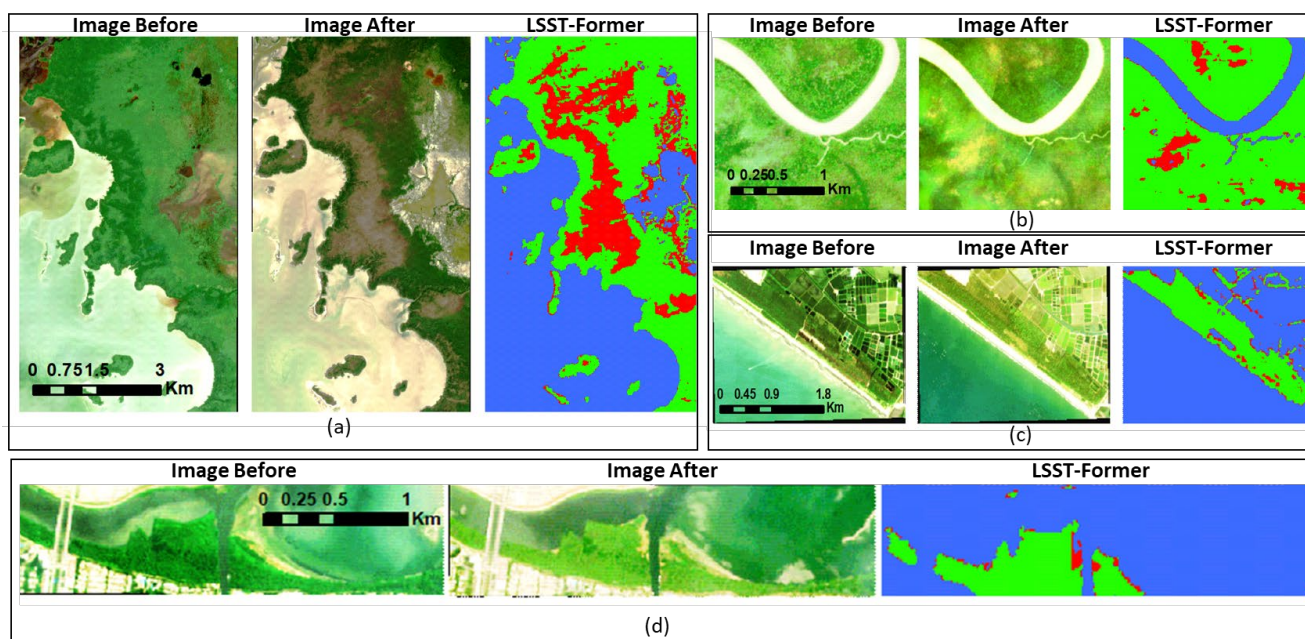
**Optimal spatial transformer encoders:** It is crucial to utilize a spatial transformer when evaluating the efficacy of the proposed method. Four distinct encoder counts (one, two, three, and four) were evaluated in order to determine which is the most efficient for a spatial transformer. SST-Former functions admirably with three and four encoders, as shown in Figure 11b. For this investigation, three encoders were chosen.

**Optimal temporal CA layers:** An investigation into the quantity of CA layers in T-Former is similarly required. Four distinct CA layer counts [1–4] were evaluated in an effort to identify the most effective number of temporal encoders. Three CA layers produce the optimal combined effect, as shown in Figure 11c.

**Loss function:** Utilizing a loss function is crucial for assessing the efficacy of the proposed method. Two different loss functions, including cross-entropy and soft cross-entropy, were evaluated to determine the most effective loss function. Figure 11d demonstrates that the SST-Former soft cross-entropy has a strong performance.

### 3.5. Universal Applicability of the Model

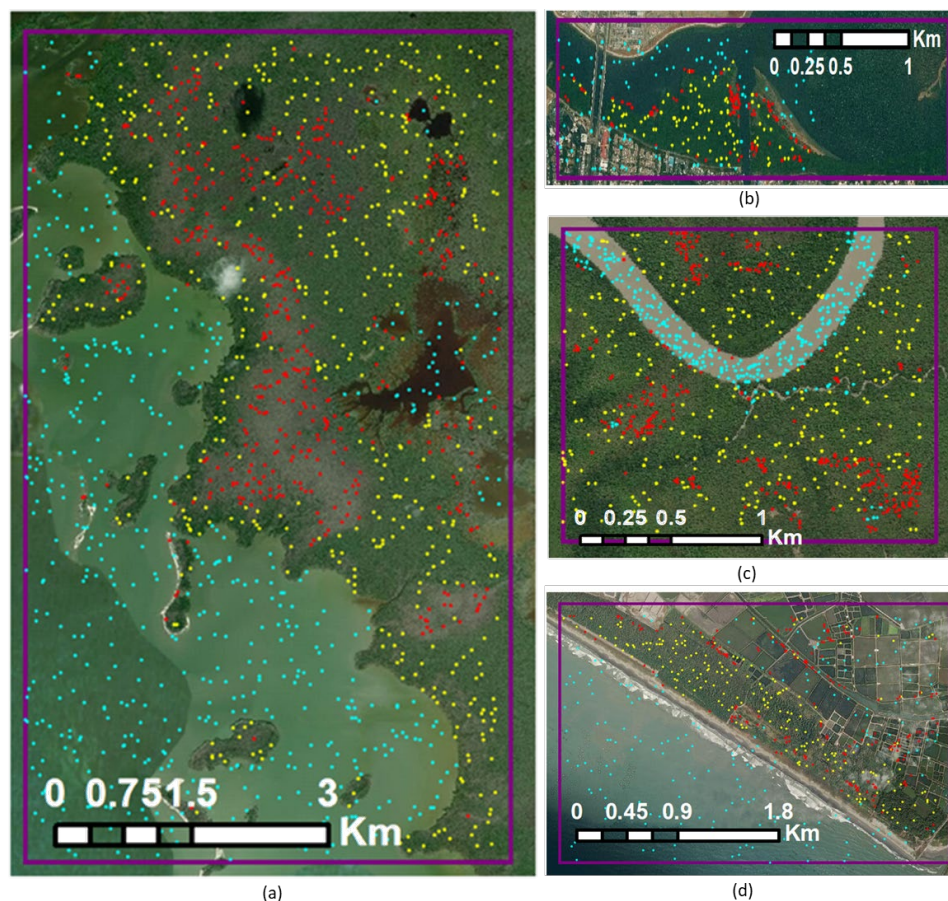
To determine whether the method is applicable to general locations, the LSST-Former model used in this study for mangrove loss detection was run in different study regions with training models in order to validate the stability of the algorithm's output. The visual results of LSST-Former are presented in Figure 12.



**Figure 12.** Universal applicability: (a) Southwest Florida-4; (b) PIK Jakarta, Indonesia; (c) Papua, Indonesia; and (d) Tainan, Taiwan. Blue, green, and red are non-mangrove, mangrove, and mangrove loss, respectively.

Prediction point samples from LSST-Former results were compared using historical ESRI World Imagery Wayback, as shown in Figure 13. A confusion matrix was employed to compute the accuracy using samples of prediction points: 1500 prediction points were validated across the Southwest Florida-4 dataset (Table 10), 300 prediction points were validated across the PIK Jakarta, Indonesia dataset (Table 11), 900 prediction points were validated across the Papua, Indonesia dataset (Table 12), and 600 prediction points were validated across the Tainan, Taiwan dataset (Table 13). Specifically, each class had 500, 100,

300, and 200 prediction points, respectively. The overall accuracy and kappa score have been calculated for map accuracy based on the confusion matrix shown in Tables 10–13. The results in Table 14 show that the overall accuracy is more than 90%, and the kappa accuracy is more than 89%. Based on these results, our model can be applied universally, even though it has a small number of labels.



**Figure 13.** Universal applicability result validation: (a) Southwest Florida-4; (b) PIK Jakarta, Indonesia; (c) Papua, Indonesia; and (d) Tainan, Taiwan. Aqua, yellow, and red are non-mangrove, mangrove, and mangrove loss, respectively.

**Table 10.** Confusion matrix for Southwest Florida-4.

Prediction	Reference Data		
	NonMg	Mg	MgLs
NonMg	493	2	5
Mg	1	492	7
MgLs	3	12	485

**Table 11.** Confusion matrix for PIK Jakarta, Indonesia.

Prediction	Reference Data		
	NonMg	Mg	MgLs
NonMg	97	3	0
Mg	0	98	2
MgLs	2	10	88

**Table 12.** Confusion matrix for Papua, Indonesia.

Prediction	Reference Data		
	NonMg	Mg	MgLs
NonMg	281	12	7
Mg	1	294	5
MgLs	5	28	267

**Table 13.** Confusion matrix for Tainan, Taiwan.

Prediction	Reference Data		
	NonMg	Mg	MgLs
NonMg	197	2	1
Mg	6	187	7
MgLs	25	2	173

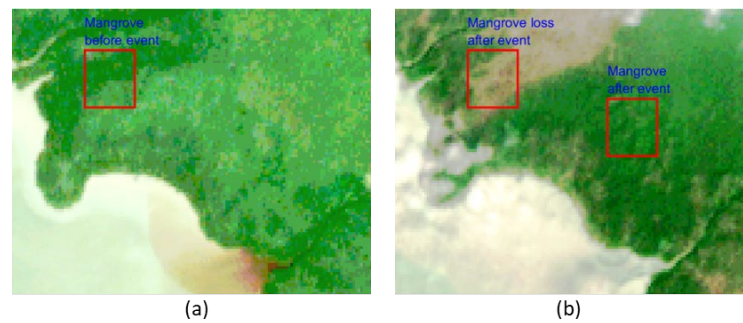
**Table 14.** Validation accuracy metrics.

	NonMg	Mg	MgLs	OA	Kappa
Southwest Florida-4	0.986	0.984	0.97	0.98	0.97
PIK Jakarta, Indonesia	0.97	0.98	0.88	0.9433	0.9150
Papua, Indonesia	0.937	0.98	0.89	0.9356	0.9033
Tainan, Taiwan	0.985	0.935	0.865	0.9283	0.8925

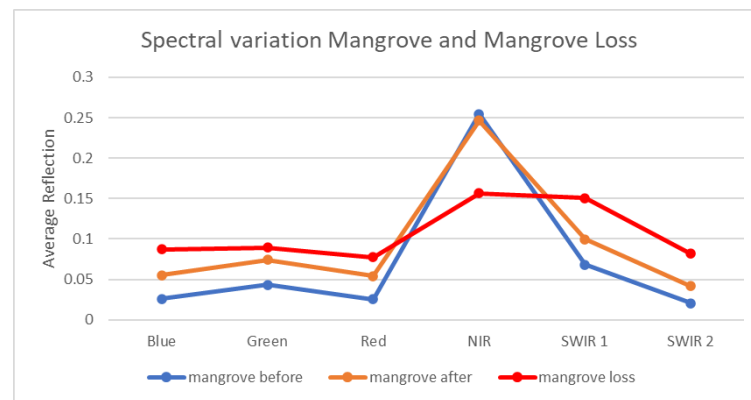
#### 4. Discussion

Integrating remote-sensing satellite images using deep learning or machine learning can increase efficiency and cost-effectiveness compared to conventional digitization approaches. We only need a small sample to use as training data for the model we have designed. Even with a small sample, we can apply the model universally. This research article presents mangrove loss detection utilizing Sentinel-2 data and a deep-learning model. The model takes into account the spatial–spectral–temporal relationship between images captured prior to and subsequent to a loss event, and the model is applied for universal applicability. As we mentioned before, the temporal resolution of satellite imagery is the main advantage that can be used for mangrove loss mapping purposes. Our model considers temporal images that take advantage of satellite imagery, which has data from the past. The proposed model with SST-Former that uses temporal images improved the mapping accuracy of mangrove loss in our study areas (Tables 7 and 8). On the other hand, the proposed model also successfully produced large intact areas and a lost mangrove map within several study areas (Figure 12). The produced large intact areas and lost mangrove maps also have satisfactory results based on the map accuracy assessments (Tables 10–14). This finding shows the applicability of remote-sensing images with the proposed deep-learning model for mangrove loss mapping in large areas and can reduce the costs required compared to direct field activities or visual interpretation.

Remote-sensing satellite images are obtained through sensors that capture electromagnetic radiation in various spectral bands. The interaction of light with the Earth’s surface provides valuable information about various features, such as land cover, vegetation health, and water bodies, causing the spectral and spatial features in satellite images to change over time and space. Figures 14 and 15 show spectral variation in images from the Southwest Florida-1 dataset before and after a hurricane event.



**Figure 14.** Southwest Florida-1 spectral location images: (a) mangrove before hurricane event and (b) mangrove loss after hurricane event and mangrove after hurricane event.



**Figure 15.** Spectral variation in mangroves and mangrove loss in Southwest Florida-1 images.

A comparison of the spectral curves of the mangrove classes before and after the hurricane shows similarities in the pattern of the spectral response. However, the spectral response of the mangrove object after the hurricane differs, with near-infrared being lower than the mangrove object; a plant with more chlorophyll will reflect more near-infrared energy than an unhealthy plant [79], and another band after the hurricane is higher than the mangrove object before the hurricane. Most mangrove objects with a low spectral response have been degraded and are classified as mangrove loss in the after-hurricane image. The SWIR bands are sensitive to the water content; it can be seen that the mangrove loss class has higher spectral reflectance in SWIR bands than the mangrove object class does.

This study considers two kinds of satellite imagery for mangrove loss detection. We took images before and after the mangrove area was lost. We assume the mangrove object exists in the image before and is then lost or degraded in the image after. The proposed LSST-Former model combines FCN and transformer architecture. The LSST-Former model considers the relationship between before and after mangrove loss using the SST-Former part. Previous studies have used the SST-Former model for hyperspectral change detection [51].

Our proposed model achieved good accuracy metrics. Our proposed model has an overall accuracy, F1-Score, and mean IoU score of 99.59%, 99.41%, and 98.84%, respectively, and IoU mangrove loss was obviously high at 97.59%. We evaluated LSST-Former by comparing it to various established architectures, including random forest (RF), Support Vector Machine (SVM), U-Net [56], LinkNet [58], Vision Transformer (ViT) [51], Spectral-Former [38], MDPrePost-Net [41], and SST-Former [53]. In the testing comparison, RF and SVM show that the architecture does not accurately detect objects; for the transformer networks, these give a better performance than other architectures, because the transformer pays attention to obtaining similarity between classes while the model considers spatial-spectral-temporal image relationships for detecting changing areas.

We investigated the training set size by examining a broad range of training sample sizes, ranging from a small sample size of 717 to a large training sample size of 2870. The effects of the various total numbers on the experimental results demonstrate a progressive rise in the overall accuracy, F1-Score, and IoU scores as the total number of the training size increases. Our method consistently achieves good accuracy—more than 90% in terms of the overall accuracy, F1-Score, and mean IoU score—even though there are only 717 labeled samples.

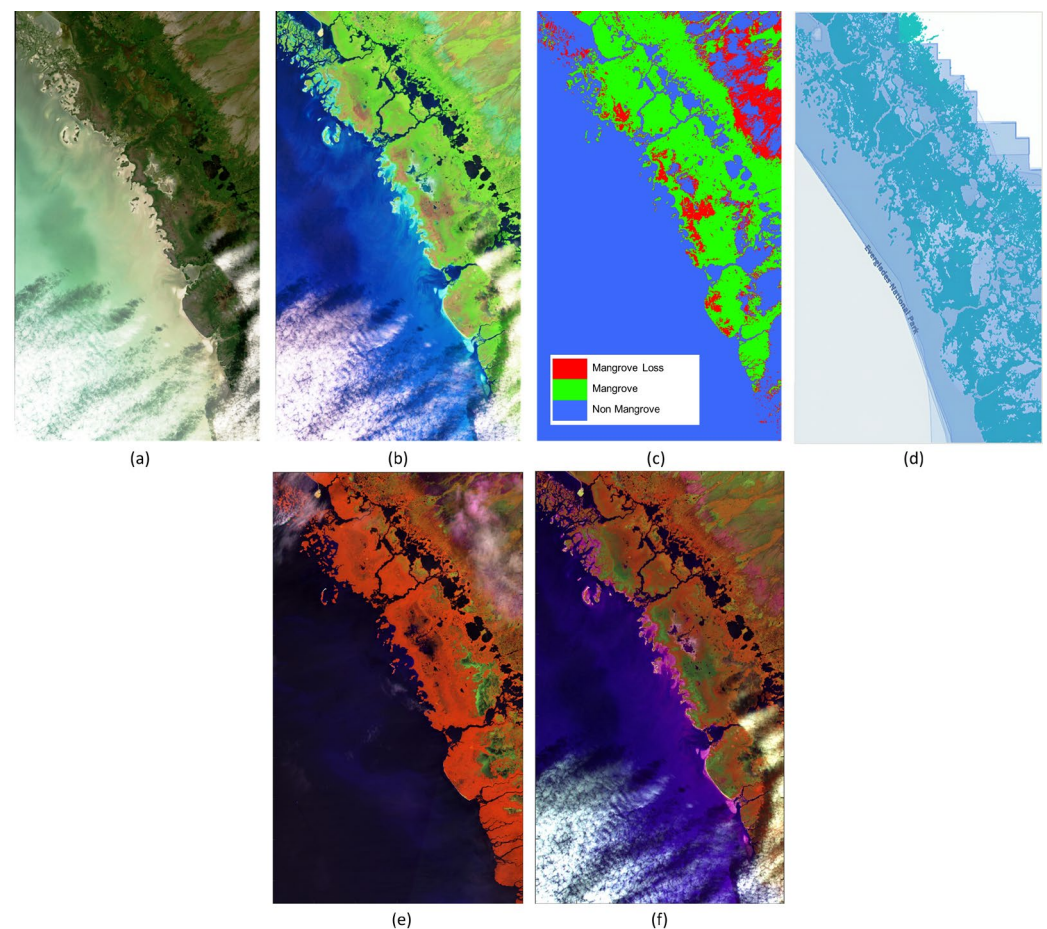
According to a prior study, the variation in input bands impacts the accuracy outcome of mangrove categorization [41]. We examined the impact of vegetation and mangrove indices on the accuracy of the results, as mangroves possess distinct spectral properties and are a distinct type of vegetation with wet vegetation features. The short-wave infrared (SWIR) band is valuable for differentiating moist objects. Several mangrove indices are established based on the spectral features specific to mangroves. Sentinel-2 can generate indices for vegetation and mangroves. The vegetation index, consisting of the CMRI, NDMI, and MMRI, exhibits a strong ability to differentiate between mangrove objects [41]. The initial dataset for this investigation consisted of 10 input bands, including blue, green, red, NIR, SWIR1, SWIR2, NDVI, CMRI, NDMI, and MMRI, which included mangrove and vegetation indices. The findings in Section 3.3 demonstrated that the incorporation of vegetation and mangrove indices significantly enhanced the categorization accuracy in a progressive way.

We evaluated the optimal parameters for the LSST-Former model, including the number of spectral encoders, the number of spatial encoders, the number of cross-attention layers, and the loss function. We found that the optimal number of spectral encoders is two, the optimal number of spatial encoders is three, the optimal number of cross-attention layers is three, and the optimal loss function is soft cross-entropy in our study.

Our result shows that the model can be applied universally and clearly distinguish between intact/healthy mangroves, mangrove loss, and non-mangrove areas. By analyzing the outcome, we may observe the distinct distribution of mangrove loss and intact mangrove areas. We validated the results with historical high-resolution images. Due to the lack of coverage of mangrove loss among the four regions, we focused on large-scale mangrove loss detection in southwest Florida, covering an area of 158,332.76 ha. A previous study used CODE-MM [42], Clark Labs, and MDPrePost-Net [41] with high-resolution Google Earth images for validation. In the present study, we used high-resolution Google Earth images, false colors, and a global mangrove watch map to validate the predictions of mangrove loss, as shown in Figure 16.

In Figure 16e, a false color composite (NIR, SWIR1, red) shows the area before the hurricane, and mangrove vegetation is displayed in orange in the composite results. Mangroves experiencing loss can be seen if we compare this with Figure 16f, with a false color composite (NIR, SWIR1, red) after the hurricane, where the mangrove, which was initially orange, experienced a change in spectral value so that it changed color in composite Figure 16f. This indicates the degradation or loss of mangrove vegetation after the hurricane. We found that 126.1852 ha of mangroves have been lost, and 391.8897 ha of mangroves remained healthy.

The results show that the overall accuracy and kappa accuracy for all study areas is more than 90%, except for kappa accuracy for the area in Tainan, Taiwan. Our validation data showed the image in high resolution, and Sentinel-2 images were quite tricky to obtain on the same date and time, because Sentinel-2 has a problem with cloud cover, and ESRI World Imagery Wayback or Google Earth images do not provide the same temporal data as the Sentinel-2 images, so we just took the closer time for validation. Sentinel-2 provides optical satellite images that are accessible but unable to penetrate clouds. We cannot use Sentinel-2 for observing objects that are covered by clouds. Data selection and location are based on the percentage of cloud cover to obtain effective results. The use of images from active remote sensing to penetrate clouds needs more attention in future research.



**Figure 16.** Large-scale prediction: (a) true-color (RGB) Google Earth image 2018/01 (after hurricane); (b) after hurricane, false color (SWIR1, NIR, red); (c) prediction; (d) Global Mangrove Watch Map 2018; (e) before hurricane, false color (NIR, SWIR1, red); and (f) after hurricane, false color (NIR, SWIR1, red).

## 5. Conclusions

This paper introduces LSST-Former, a novel deep-learning network that utilizes few-shot learning to identify mangrove loss in a small number of labeled data. Our approach effectively captures spatial, spectral, and temporal information from Sentinel-2 images while maintaining a relatively simple model structure using a combination of FCN and a transformer network, even with limited labeled samples. Furthermore, our approach was compared to many other approaches. The experimental results indicate that the LSST-Former approach outperforms other methods in accuracy and resilience for detecting mangrove loss using Sentinel-2 images. Finally, experiments were carried out on four distinct regions, and our approach was compared to a high-resolution image. The experimental results indicate that the LSST-Former technique exhibits superior accuracy and resilience in detecting mangrove loss.

**Author Contributions:** Methodology, I.A.P. and Y.-N.C.; Validation, I.A.P., I.J. and S.-N.L.; Resources, Y.-N.C.; Data curation, S.-N.L.; Writing—original draft, I.A.P.; Writing—review & editing, I.J. and Y.-N.C.; Supervision, Y.-N.C.; Project administration, Y.-N.C. and K.-C.F.; Funding acquisition, Y.-N.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The Sentinel-2 images are available at the URL: <https://developers.google.com/earth-engine/datasets/catalog/sentinel-2> (accessed on 4 November 2023). While visually interpreted labels and derived mangrove degradation map are available by the request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Metternicht, G.; Lucas, R.; Bunting, P.; Held, A.; Lyburner, L.; Ticehurst, C. Addressing Mangrove Protection in Australia: The Contribution of Earth Observation Technologies. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 6548–6551.
2. Sidik, F.; Supriyanto, B.; Krisnawati, H.; Muttaqin, M.Z. Mangrove Conservation for Climate Change Mitigation in Indonesia. *WIREs Clim. Chang.* **2018**, *9*, e529. [[CrossRef](#)]
3. Chow, J. Mangrove management for climate change adaptation and sustainable development in coastal zones. *J. Sustain. For.* **2018**, *37*, 139–156. [[CrossRef](#)]
4. Islam, M.D.; Di, L.; Mia, M.R.; Sithi, M.S. Deforestation Mapping of Sundarbans Using Multi-Temporal Sentinel-2 Data & Transfer Learning. In Proceedings of the 10th International Conference on Agro-Geoinformatics (Agro-Geoinformatics) 2022, Quebec City, QC, Canada, 11–14 July 2022.
5. Arifanti, V.B.; Sidik, F.; Mulyanto, B.; Susilowati, A.; Wahyuni, T.; Subarno, S.; Yulianti, Y.; Yuniarti, N.; Aminah, A.; Suita, E. Challenges and Strategies for Sustainable Mangrove Management in Indonesia: A Review. *Forests* **2022**, *13*, 695. [[CrossRef](#)]
6. Wong, W.Y.; Al-Ani, A.K.I.; Hasikin, K.; Khairuddin, A.S.M.; Razak, S.A.; Hizaddin, H.F.; Mokhtar, M.I.; Azizan, M.M. Water, Soil and Air Pollutants' Interaction on Mangrove Ecosystem and Corresponding Artificial Intelligence Techniques Used in Decision Support Systems—A Review. *IEEE Access* **2021**, *9*, 105532–105563. [[CrossRef](#)]
7. Sunkur, R.; Kantamaneni, K.; Bokhoree, C.; Ravan, S. Mangroves' Role in Supporting Ecosystem-Based Techniques to Reduce Disaster Risk and Adapt to Climate Change: A Review. *J. Sea Res.* **2023**, *196*, 102449. [[CrossRef](#)]
8. Trégarot, E.; Caillaud, A.; Cornet, C.C.; Taureau, F.; Catty, T.; Cragg, S.M.; Failler, P. Mangrove Ecological Services at the Forefront of Coastal Change in the French Overseas Territories. *Sci. Total Environ.* **2021**, *763*, 143004. [[CrossRef](#)]
9. Gomes, L.E.d.O.; Sanders, C.J.; Nobrega, G.N.; Vescovi, L.C.; Queiroz, H.M.; Kauffman, J.B.; Ferreira, T.O.; Bernardino, A.F. Ecosystem Carbon Losses Following a Climate-Induced Mangrove Mortality in Brazil. *J. Environ. Manag.* **2021**, *297*, 113381. [[CrossRef](#)] [[PubMed](#)]
10. Ward, R.D.; Drude de Lacerda, L. Responses of mangrove ecosystems to sea level change. In *Dynamic Sedimentary Environments of Mangrove Coasts*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 235–253.
11. Vizcaya-Martínez, D.A.; Flores-de-Santiago, F.; Valderrama-Landeros, L.; Serrano, D.; Rodríguez-Sobreyra, R.; Álvarez-Sánchez, L.F.; Flores-Verdugo, F. Monitoring Detailed Mangrove Hurricane Damage and Early Recovery Using Multisource Remote Sensing Data. *J. Environ. Manag.* **2022**, *320*, 115830. [[CrossRef](#)] [[PubMed](#)]
12. Kudrass, H.R.; Hanebuth, T.J.J.; Zander, A.M.; Linstädter, J.; Akther, S.H.; Shohrab, U.M. Architecture and Function of Salt-Producing Kilns from the 8th to 18th Century in the Coastal Sundarbans Mangrove Forest, Central Ganges-Brahmaputra Delta, Bangladesh. *Archaeol. Res. Asia* **2022**, *32*, 100412. [[CrossRef](#)]
13. Chopade, M.R.; Mahajan, S.; Chaube, N. Assessment of land use, land cover change in the mangrove forest of Ghogha area, Gulf of Khambhat, Gujarat. *Expert Syst. Appl.* **2023**, *212*, 118839. [[CrossRef](#)]
14. Quevedo, J.M.D.; Lukman, K.M.; Ulumuddin, Y.I.; Uchiyama, Y.; Kohsaka, R. Applying the DPSIR Framework to Qualitatively Assess the Globally Important Mangrove Ecosystems of Indonesia: A Review towards Evidence-Based Policymaking Approaches. *Mar. Policy* **2023**, *147*, 105354. [[CrossRef](#)]
15. Gitau, P.N.; Duvail, S.; Verschuren, D. Evaluating the combined impacts of hydrological change, coastal dynamics and human activity on mangrove cover and health in the Tana River delta, Kenya. *Reg. Stud. Mar. Sci.* **2023**, *61*, 102898. [[CrossRef](#)]
16. Numere, A.O. Impact of anthropogenic activities on mangrove forest health in urban areas of the Niger Delta: Its susceptibility and sustainability. In *Water, Land, and Forest Susceptibility and Sustainability*; Academic Press: Cambridge, MA, USA, 2023; pp. 459–480.
17. Long, C.; Dai, Z.; Zhou, X.; Mei, X.; Mai Van, C. Mapping Mangrove Forests in the Red River Delta, Vietnam. *For. Ecol. Manag.* **2021**, *483*, 118910. [[CrossRef](#)]
18. Valiela, I.; Bowen, J.L.; York, J.K. Mangrove Forests: One of the World's Threatened Major Tropical Environments. *Bioscience* **2001**, *51*, 807–815. [[CrossRef](#)]
19. Goldberg, L.; Lagomasino, D.; Thomas, N.; Fatoyinbo, T. Global Declines in Human-driven Mangrove Loss. *Glob. Chang. Biol.* **2020**, *26*, 5844–5855. [[CrossRef](#)]
20. Hamilton, S.E.; Casey, D. Creation of a High Spatio-Temporal Resolution Global Database of Continuous Mangrove Forest Cover for the 21st Century (CGMFC-21). *Glob. Ecol. Biogeogr.* **2016**, *25*, 729–738. [[CrossRef](#)]
21. Zhang, R.; Jia, M.; Wang, Z.; Zhou, Y.; Wen, X.; Tan, Y.; Cheng, L. A comparison of Gaofen-2 and Sentinel-2 imagery for mapping mangrove forests using object-oriented analysis and random forest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4185–4193. [[CrossRef](#)]
22. Chen, Z.; Zhang, M.; Zhang, H.; Liu, Y. Mapping Mangrove Using a Red-Edge Mangrove Index (REMI) Based on Sentinel-2 Multispectral Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4409511. [[CrossRef](#)]
23. Zhu, Y.; Liu, K.; Liu, L.; Myint, S.W.; Wang, S.; Cao, J.; Wu, Z. Estimating and Mapping Mangrove Biomass Dynamic Change Using WorldView-2 Images and Digital Surface Models. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2123–2134. [[CrossRef](#)]

24. Xue, Z.; Qian, S. Generalized composite mangrove index for mapping mangroves using Sentinel-2 time series data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 5131–5146. [[CrossRef](#)]
25. Magris, R.A.; Barreto, R. Mapping and assessment of protection of mangrove habitats in Brazil. *Panam. J. Aquat. Sci.* **2010**, *5*, 546–556.
26. Pham, T.D.; Yokoya, N.; Bui, D.T.; Yoshino, K.; Friess, D.A. Remote Sensing Approaches for Monitoring Mangrove Species, Structure, and Biomass: Opportunities and Challenges. *Remote Sens.* **2019**, *11*, 230. [[CrossRef](#)]
27. Yu, C.; Liu, B.; Deng, S.; Li, Z.; Liu, W.; Ye, D.; Hu, J.; Peng, X. Using Medium-Resolution Remote Sensing Satellite Images to Evaluate Recent Changes and Future Development Trends of Mangrove Forests on Hainan Island, China. *Forests* **2023**, *14*, 2217. [[CrossRef](#)]
28. Hu, T.; Zhang, Y.; Su, Y.; Zheng, Y.; Lin, G.; Guo, Q. Mapping the Global Mangrove Forest Aboveground Biomass Using Multisource Remote Sensing Data. *Remote Sens.* **2020**, *12*, 1690. [[CrossRef](#)]
29. Zhao, C.P.; Qin, C.Z. A detailed mangrove map of China for 2019 derived from Sentinel-1 and-2 images and Google Earth images. *Geosci. Data J.* **2022**, *9*, 74–88. [[CrossRef](#)]
30. Sharifi, A.; Felegari, S.; Tariq, A. Mangrove forests mapping using Sentinel-1 and Sentinel-2 satellite images. *Arab. J. Geosci.* **2022**, *15*, 1593. [[CrossRef](#)]
31. Giri, C.; Pengra, B.; Long, J.; Loveland, T.R. Next Generation of Global Land Cover Characterization, Mapping, and Monitoring. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *25*, 30–37. [[CrossRef](#)]
32. Rijal, S.S.; Pham, T.D.; Noer'Aulia, S.; Putera, M.I.; Saintilan, N. Mapping Mangrove Above-Ground Carbon Using Multi-Source Remote Sensing Data and Machine Learning Approach in Loh Buaya, Komodo National Park, Indonesia. *Forests* **2023**, *14*, 94. [[CrossRef](#)]
33. Soltanikazemi, M.; Minaei, S.; Shafizadeh-Moghadam, H.; Mahdavian, A. Field-Scale Estimation of Sugarcane Leaf Nitrogen Content Using Vegetation Indices and Spectral Bands of Sentinel-2: Application of Random Forest and Support Vector Regression. *Comput. Electron. Agric.* **2022**, *200*, 107130. [[CrossRef](#)]
34. Xu, C.; Wang, J.; Sang, Y.; Li, K.; Liu, J.; Yang, G. An Effective Deep Learning Model for Monitoring Mangroves: A Case Study of the Indus Delta. *Remote Sens.* **2023**, *15*, 2220. [[CrossRef](#)]
35. Guo, Y.; Liao, J.; Shen, G. Mapping large-scale mangroves along the maritime silk road from 1990 to 2015 using a novel deep learning model and landsat data. *Remote Sens.* **2021**, *13*, 245. [[CrossRef](#)]
36. Wang, Y.; Gu, L.; Jiang, T.; Gao, F. MDE-U-Net: A Multitask Deformable U-Net Combined Enhancement Network for Farmland Boundary Segmentation. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 3001305.
37. Tran, T.L.C.; Huang, Z.C.; Tseng, K.H.; Chou, P.H. Detection of Bottle Marine Debris Using Unmanned Aerial Vehicles and Machine Learning Techniques. *Drones* **2022**, *6*, 401. [[CrossRef](#)]
38. Hong, D.; Han, Z.; Yao, J.; Gao, L.; Zhang, B.; Plaza, A.; Chanussot, J. SpectralFormer: Rethinking Hyperspectral Image Classification With Transformers. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5518615. [[CrossRef](#)]
39. Jamaluddin, I.; Chen, Y.N.; Ridha, S.M.; Mahyatar, P.; Ayudyanti, A.G. Two Decades Mangroves Loss Monitoring Using Random Forest and Landsat Data in East Luwu, Indonesia (2000–2020). *Geomatics* **2022**, *2*, 282–296. [[CrossRef](#)]
40. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [[CrossRef](#)] [[PubMed](#)]
41. Jamaluddin, I.; Thaipisutikul, T.; Chen, Y.N.; Chuang, C.H.; Hu, C.L. MDPrePost-Net: A Spatial-Spectral-Temporal Fully Convolutional Network for Mapping of Mangrove Degradation Affected by Hurricane Irma 2017 Using Sentinel-2 Data. *Remote Sens.* **2021**, *13*, 5042. [[CrossRef](#)]
42. Lin, C.H.; Chu, M.C.; Tang, P.W. CODE-MM: Convex Deep Mangrove Mapping Algorithm Based On Optical Satellite Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5620619. [[CrossRef](#)]
43. Iovan, C.; Kulbicki, M.; Mermert, E. Deep convolutional neural network for mangrove mapping. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020.
44. Diniz, C.; Cortinhas, L.; Nerino, G.; Rodrigues, J.; Sadeck, L.; Adami, M.; Souza-Filho, P. Brazilian Mangrove Status: Three Decades of Satellite Data Analysis. *Remote Sens.* **2019**, *11*, 808. [[CrossRef](#)]
45. Chen, N. Mapping mangrove in Dongzhaigang, China using Sentinel-2 imagery. *J. Appl. Remote Sens.* **2020**, *14*, 014508. [[CrossRef](#)]
46. Xue, Z.; Qian, S. Two-Stream Translating LSTM Network for Mangroves Mapping Using Sentinel-2 Multivariate Time Series. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4401416. [[CrossRef](#)]
47. Al Dogom, D.W.; Samour, B.M.M.; Al Shamsi, M.; Almansoori, S.; Aburaed, N.; Zitouni, M.S. Machine Learning for Spatiotemporal Mapping and Monitoring of Mangroves and Shoreline Changes Along a Coastal Arid Region. In Proceedings of the IGARSS 2023—2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023.
48. Li, L.; Zhang, W.; Zhang, X.; Emam, M.; Jing, W. Semi-Supervised Remote Sensing Image Semantic Segmentation Method Based on Deep Learning. *Electronics* **2023**, *12*, 348. [[CrossRef](#)]
49. Wan, Q.; Ji, H.; Shen, L. Self-attention based text knowledge mining for text detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 5979–5988.
50. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. In Proceedings of the Computer Vision—ECCV 2020, Glasgow, UK, 23–28 August 2020; pp. 213–229.
51. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.



52. Sun, L.; Zhao, G.; Zheng, Y.; Wu, Z. Spectral–Spatial Feature Tokenization Transformer for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5522214. [[CrossRef](#)]
53. Wang, Y.; Hong, D.; Sha, J.; Gao, L.; Liu, L.; Zhang, Y.; Rong, X. Spectral–Spatial–Temporal Transformers for Hyperspectral Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5536814. [[CrossRef](#)]
54. Liu, Y.; Hu, J.; Kang, X.; Luo, J.; Fan, S. Interactformer: Interactive Transformer and CNN for Hyperspectral Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5531715. [[CrossRef](#)]
55. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [[CrossRef](#)] [[PubMed](#)]
56. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, Munich, Germany, 5–9 October 2015; pp. 234–241.
57. Chaurasia, A.; Culurciello, E. LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017.
58. Saferbekov, S.; Igloukov, V.; Buslaev, A.; Shvets, A. Feature Pyramid Network for Multi-Class Land Segmentation. *Comput. Vis. Pattern Recognit.* **2018**, *2*, 272–2723.
59. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
60. Ren, M.; Triantafillou, E.; Ravi, S.; Snell, J.; Swersky, K.; Tenenbaum, J.; Larochelle, H.; Zemel, R. Meta-Learning for Semi-Supervised Few-Shot Classification. *arXiv* **2018**, arXiv:1803.00676.
61. Liu, B.; Gao, K.; Yu, A.; Ding, L.; Qiu, C.; Li, J. ES2FL: Ensemble Self-Supervised Feature Learning for Small Sample Classification of Hyperspectral Images. *Remote Sens.* **2022**, *14*, 4236. [[CrossRef](#)]
62. Hamilton, S.E.; Friess, D.A. Global Carbon Stocks and Potential Emissions Due to Mangrove Deforestation from 2000 to 2012. *Nat. Clim. Chang.* **2018**, *8*, 240–244. [[CrossRef](#)]
63. Rumondang, A.L.; Kusmana, C.; Budi, S.W. Species Composition and Structure of Angke Kapuk Mangrove Protected Forest, Jakarta, Indonesia. *Biodiversitas J. Biol. Divers.* **2021**, *22*, 9. [[CrossRef](#)]
64. Liu, C.C.; Hsu, T.W.; Wen, H.L.; Wang, K.H. Mapping Pure Mangrove Patches in Small Corridors and Sandbanks Using Airborne Hyperspectral Imagery. *Remote Sens.* **2019**, *11*, 592. [[CrossRef](#)]
65. Google Earth Engine. 2023. Available online: <https://developers.google.com/earth-engine/datasets/catalog/sentinel-2> (accessed on 4 November 2023).
66. Yin, F.; Lewis, P.E.; Gómez-Dans, J.L. Bayesian Atmospheric Correction over Land: Sentinel-2/MSI and Landsat 8/OLI. 2022. Available online: <https://gmd.copernicus.org/articles/15/7933/2022/> (accessed on 11 March 2024).
67. Shi, T.; Liu, J.; Hu, Z.; Liu, H.; Wang, J.; Wu, G. New spectral metrics for mangrove forest identification. *Remote Sens. Lett.* **2016**, *7*, 885–894. [[CrossRef](#)]
68. Gupta, K.; Mukhopadhyay, A.; Giri, S.; Chanda, A.; Datta Majumdar, S.; Samanta, S.; Mitra, D.; Samal, R.N.; Pattnaik, A.K.; Hazra, S. An index for discrimination of mangroves from non-mangroves using LANDSAT 8 OLI imagery. *MethodsX* **2018**, *5*, 1129–1139. [[CrossRef](#)]
69. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [[CrossRef](#)]
70. McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [[CrossRef](#)]
71. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [[CrossRef](#)]
72. Nanduri, A.; Chellappa, R. Semi-Supervised Cross-Spectral Face Recognition With Small Datasets. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2024.
73. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016.
74. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
75. Gehring, J.; Auli, M.; Grangier, D.; Yarats, D.; Dauphin, Y.N. Convolutional sequence to sequence learning. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1243–1252.
76. Rezatofighi, S.H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.D.; Savarese, S. Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
77. Sokolova, M.; Japkowicz, N.; Szpakowicz, S. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. *AI 2006 Adv. Artif. Intell.* **2006**, *4304*, 1015–1021.
78. ESRI | World Imagery Wayback. 2023. Available online: <https://livingatlas.arcgis.com/wayback/> (accessed on 8 December 2023).
79. Reflected Near-Infrared Waves. Available online: [https://science.nasa.gov/ems/08\\_nearinfraredwaves/](https://science.nasa.gov/ems/08_nearinfraredwaves/) (accessed on 9 February 2024).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.