Vietnam Journal of Earth Sciences, 47(2), 272–289, https://doi.org/10.15625/2615-9783/23010

Vietnam Journal of Earth Sciences
https://vjs.ac.vn/index.php/jse

Observation of coal mining activities in Ha Long and Cam Pha cities, Quang Ninh province, Vietnam using Sentinel-2 data

Dang Vu Khac[1], Ta Duc Hieu[2,*], Ngô Thị Hải Yến[1], Nguyen Khac Anh[1], Phạm Thanh Hải[3], Christiane Weber[4]

[1]*Faculty of Geography, Hanoi National University of Education, Vietnam*
[2]*Faculty of Geography, Ho Chi Minh City University of Education, Vietnam*
[3]*Institute of Earth Sciences, Vietnam Academy of Sciences and Technology, Vietnam*
[4]*UMR TETIS, CNRS, Université de Montpellier, France*

Received April 04 2025; Received in revised form 15 May 2025; Accepted 09 June 2025

ABSTRACT

Mapping the spatial distribution of exposed coal sites over a large territory and ensuring temporal continuity is essential for environmental assessment. It helps identify relationships between coal mining activities and environmental issues, such as water pollution from coal contamination and air pollution from particulate matter in the coastal areas of Ha Long and Cam Pha. The carbon index has proven to be an effortless yet effective technique. It is calculated using the spectral ratio between two shortwave infrared bands (Band 11 - SWIR1 and Band 12 - SWIR2) of Sentinel-2 imagery. The selection of an appropriate carbon threshold value - used to distinguish exposed coal from other land cover types, particularly urban areas - was validated through visual interpretation of high-resolution satellite images from the Google Earth dataset. This approach, which leverages Sentinel-2 imagery with higher spatial resolution, differs from previous studies that employed complex algorithms or multiple spectral indicators to detect exposed coal sites using lower-resolution images. The results, with an overall accuracy of 87.35% and a Kappa coefficient of 0.73, provide valuable support for monitoring and managing the exploitation of natural resources. Additionally, this method identifies illegal coal mines and seawater contamination, aiding in the management of the coal industry, among other benefits. These findings lay a crucial foundation for assessing the environmental impact of coal mining activities and proposing solutions for ecological restoration in coastal areas where diverse and complex socio-economic activities occur.

Keywords: Carbon index, exposed coal site, Sentinel-2 image, visual interpretation, thresholding classification.

1. Introduction

The coastal area of Ha Long-Cam Pha is home to the largest concentration of coal mining operations and logistics services in northeast Vietnam. This region contains several coal mines with substantial reserves. Despite efforts by local authorities to mitigate the environmental impact of the mining industry, the combined effects of coal mining activities and rapid socio-economic development in recent years have significantly altered the coastal ecosystems. These changes directly impact the environmental quality in Ha Long and Cam Pha cities, primarily as many coal mines are near densely populated settlements. For example, opencast coal

*Corresponding author, Email: hieutd@hcmue.edu.vn

mining removes the rock cover, lowers the ground surface (Nguyen, 2015), alters land cover and land use (Hens et al., 2000), disrupts drainage networks (Mustafin, Tran, & Tran, 2019); and releases dust into the air due to the transportation of waste rock and raw coal, etc. (Pham, Do, Pham, Phan, & Pham, 2020; Tran, Cao, Vu, Tran, & Tran, 2012).

Exposed coal sites are among the sources that can contribute to environmental problems. They refer to areas where raw coal, extracted through opencast or underground mining, is exposed to outdoor conditions. Exposed coal sites comprise extraction sites (mining areas) and coal yard sites (storage areas), which continually change as mining operations progress. Additionally, acceptable debris from exposed coal forms coal slurry, which is carried into drainage networks and transported to the sea during the rainy season. This leads to seawater contamination in Bai Tu Long Bay and Ha Long Bay (Cao et al., 2020), both of which are UNESCO World Heritage sites recognized for their unique geological and geomorphological features.

Furthermore, illegal coal mining remains a persistent issue due to the region's widespread distribution of coal seams, resulting in the loss of public resources (Van, 2015). Therefore, identifying exposed coal sites across a large territory and ensuring temporal continuity is crucial for generating fundamental data to establish links with ongoing environmental events. The results will be a reliable foundation for supporting ecological and natural resource management in this coastal region.

Raw coal is exposed at coalfields during mining activities or accumulated at coal yard sites after mining. Measuring the reflectance value of raw coal at several wavelengths in the spectrum helps distinguish it from other types of land cover through a radiometric curve (Fig. 1). Satellite imagery can meet this need, offering significant benefits when combined with ground surveys (Mao et al., 2014; E. Yang, Ge, & Wang, 2018). Moreover, with the increasing availability of satellite images, they should be considered a practical resource for terrestrial observation, especially since Sentinel-2 data are freely accessible to all users. Therefore, promoting the use of these data to locate exposed coal sites, including extraction sites and coal yard sites, presents valuable opportunities for environmental monitoring and natural resource protection under the specific conditions of Vietnam.

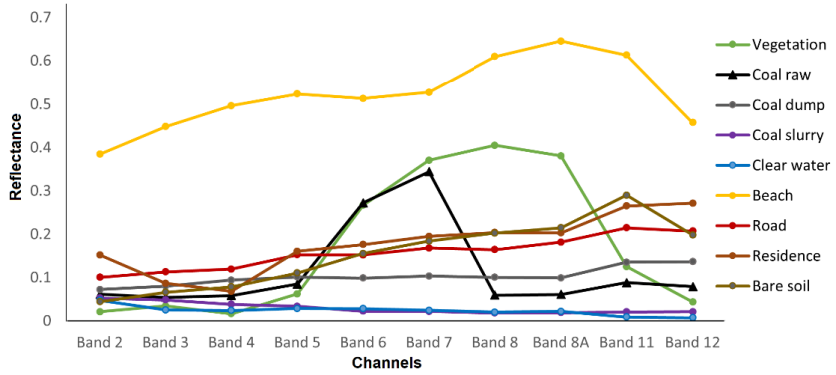


Figure 1. Spectral profiles of Sentinel-2 imagery for the major land cover types in the study area

Over the past few decades, image processing techniques for land cover mapping have undergone significant development worldwide. Remote sensing has several advantages, including wide coverage, short temporal cycle, and reduced time and cost

compared to traditional field survey-based methods (Giri, 2012). Numerous studies have utilized satellite imagery to identify the locations of exposed coal sites (Demirel, Emil, & Duzgun, 2010; Song, Song, Gu, & Li, 2020; Xiao et al., 2016). Generally, two main approaches are used for this purpose: In the first approach, classification based on their unique spectral patterns and spatial information involves grouping ground features into different categories. Algorithms such as maximum likelihood (Azeez & Mukhitdinov, 2020), Random Forest (Pandey, Mishra, Swamy, Anderson, & Thakur, 2025), and object-based classification (Yuan et al., 2019) have been successfully applied. Results indicate that the object-based method (Gao, Kerle, & Mas, 2009) and the Support Vector Machine method (Chen, Li, & Wang, 2020) can effectively utilize multi-spectral images with high spatial resolution to identify ground features in coal mining areas. In the second approach, coal mining areas have also been delineated using spectral ratios such as the normalized differential carbon index, derived from the Red and Near Infrared bands of the Landsat TM satellite (Mao et al., 2014) or two shortwave infrared bands (SWIR1 and SWIR2) of the Landsat 8 satellite (Mukherjee, Mukherjee, Chakravarty, & Aikat, 2018), or from the combination of these Blue, Green, Red, Near Infrared, SWIR1 and SWIR2 bands of the Landsat 8 satellite (Z. Yang, He, Zhang, & Zhao, 2024). These carbon indices enable the identification of coal mining areas with high accuracy by leveraging differences between the reflectance curve of various land cover types and the reflectance curve of raw coal, as measured using hand-held spectrophotometer. Findings from these studies can be applied to monitor coal mines, track land-use dynamics, manage wastewater and coal seams, and support environmental reclamation efforts in coal mining areas.

Despite the increasing availability of Sentinel-2 data following the launch of the two Sentinel 2A and 2B satellites by the European Space Agency (ESA) for capturing Earth surface imagery, research employing this second approach with Sentinel-2 data remains limited. Therefore, this new generation of satellite imagery is a highly suitable choice for identifying the distribution of exposed coal sites in Ha Long and Cam Pha cities due to its medium spatial resolution, high temporal frequency, and a spectral band configuration well-suited for calculating the spectral ratio. The results are compared with those derived from the visual interpretation of high-resolution images (from the Google Earth dataset) to assess accuracy.

2. Study area

The study area is primarily located in Ha Long and Cam Pha cities, Quang Ninh Province. These two adjacent cities occupy a strategic position in the Northeastern coastal region of the country. The area features remarkably diverse landscapes, with hills and mountains dominating, ranging in elevation from 0 to 980 m, alongside narrow coastal alluvial plains. The remaining portion comprises a vast marine area containing hundreds of large and small limestone islands. The combined territory of these two cities covers approximately 758 km^2, spanning from latitude 20°54'35"N to 21°14'05"N and longitude 106°54'08"E to 107°24'08"E (Fig. 2). The region experiences a tropical monsoon climate, with an annual average temperature of 23.7°C and an average yearly precipitation of approximately 1832 mm. The climate is distinctly seasonal, with a hot, rainy summer (May to October), accounting for 80% to 85% of total annual precipitation, and a cold, dry winter (November to April), contributing only 15% to 20% of total annual rainfall (Huynh, 2002). The hydrographic network primarily consists of small, short rivers originating from the mountainous districts in the West. These rivers exhibit rapid flow increases and quick drainage to the

sea due to the region's steep terrain and heavy rainfall during the rainy season.

The underground of the coastal area in Ha Long and Cam Pha cities consists of geological formations that developed from the Paleozoic era, such as the Tan Mai Formation, to the Cenozoic era, including the Ha Coi Formation. However, some formations exhibit distinctive characteristics unique to this coastal area. For instance, the Hon Gai Formation contains the most valuable coal seams, which have been industrially exploited since the late 19th century and continue to be mined today (Tran, 1998). Numerous coal mines are concentrated in the central part of the study area. At the same time, residential settlements have developed on both sides of National Road No18, with various artificial structures extending along the shoreline.

Meanwhile, the Northern part of the area remains dominated by tropical rainforest. (Fig. 2). Additionally, carbonate sediments of the Bai Chay Formation, the Bac Son Formation, and the Cat Ba Formation contain thick limestone layers formed in the marine environment, contributing to the region's unique landscapes. The United Nations Educational, Scientific and Cultural Organization (UNESCO) recognized the area's geomorphological value by designating it as a World Natural Heritage site in 1994 (Waltham, 1998). The Hon Gai formation is exposed in the coastal regions of Ha Long and Cam Pha. It consists of two parts: (1) lower part: this section has a thickness ranging from 460 to 1430 m, with a lithological composition of gravel-stone, sandstone, siltstone, coaly shale, and 19 to 23 commercially valuable coal seams; (2) upper part: this section has a thickness ranging from 400 to 500 m and consists of conglomerate, sandstone, siltstone, and coaly shale, with the presence of 1 to 5 thin coal seams (Tong et al., 2005). Since the late Ordovician period, geological activity in the Quang Ninh coal basin has been diverse and complex.

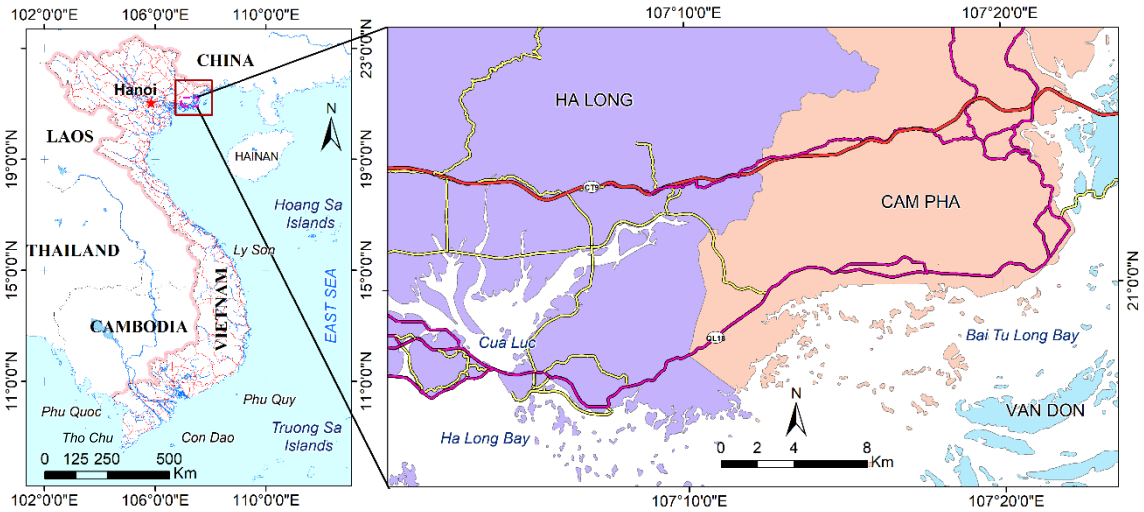


Figure 2. Map of Northern Vietnam showing the study area (pink dashed rectangle) and the Sentinel-2 scene boundary (brown rectangle)

Although no intrusive igneous rock has been detected, numerous geologic faults have fractured the rock mass along primary directions, including East-West, North-South, Northeast-Southwest, and Northwest-Southeast. Additionally, short asymmetrical folds have developed, with horizontally displaced limbs (Mai, Quyen, & Luong,

1983). These characteristics present various challenges for investigating, exploring, and mining coal seams.

The coastline extends from Dam Ha to Quang Yen and includes famous cultural and historical sites, such as Cua Ong Temple, Vung Duc Spot, and Ha Long Bay, a UNESCO World Natural Heritage site. The region is also rich in underground mineral resources, including coal, limestone, clay, and mineral water. These natural advantages provide a competitive edge for developing key economic sectors, such as industry, aquaculture, and tourism. Given its natural landscape and marine tourism potential, Ha Long and Cam Pha have been identified as model cities for transforming "*grey*" to "*green*" development, ensuring sustainable growth and environmental protection. Local authorities have planned and reorganized functional urban areas, improved urban infrastructures, expanded public facilities, and enhanced ecological sanitation. Therefore, Cam Pha was designated as a class II urban center in 2012, while Ha Long was recognized as a class I urban center in 2013. Urban classification is defined in the report by Cira et al. (By 2019, the combined population of the two cities had reached 609,980 people, distributed across 20 wards in Ha Long City and 17 wards in Cam Pha City, covering a total area of 758 km², with an average density of 805 people per km². Along with rapid urbanization, dynamic socio-economic activities generate significant environmental pollutants, such as wastewater, solid waste, and exhaust gas, which must be treated to comply with legal regulations. However, local authorities face significant challenges in environmental management due to limited capacity and the complexity of socio-economic activities, which are closely intertwined with the urbanization process in a small territory (NIURP, 2012).

3. Method and data

3.1. Data

3.1.1. Sentinel-2 imagery

The selected image was taken in late autumn 2019 under favorable weather conditions, with no cloud cover over the study area. As a result, the ground features are visible in the image, whose parameters are presented in Table 1 below.

Table 1. Sentinel-2 image parameters used in this study

Nomenclature	**S2B_MSIL2A_20191212T032129_N0500_R118_T48QYJ_20230804T102754**
Satellite	Sentinel-2B
Sensor	MSI
Processing level	2A
Acquisition date	12/12/2019
Cloud cover	2.14 %
Orbit number	118
Direction	Descending

Sentinel-2 image was downloaded from the European Space Agency (ESA) website: https://scihub.copernicus.eu/. The image was preprocessed at level 2A, providing the Bottom Of Atmosphere (BOA) reflectance for each pixel, and was geo-referenced using the UTM coordinate system (WGS84). Sentinel-2 captures 13 spectral bands in the visible and infrared regions of the spectrum, with different spatial resolutions: 10 m, 20 m, and 60 m (Table 2). In this study, two shortwave infrared bands (SWIR1 and SWIR2), with a spatial resolution of 20 m, were used to identify coal extraction sites and yards.

Table 2. Sentinel-2 band characteristics (ESA, 2021a)

Bands	Central wavelength (nm)	Spatial resolution (m)
Band 1 - Coastal aerosol	442.3	60
Band 2 - Blue	492.1	10
Band 3 - Green	559.0	10
Band 4- Red	665.0	10
Band 5 - Vegetation red edge 1	703.8	20
Band 6 - Vegetation red edge 2	739.1	20
Band 7 - Vegetation red edge 3	779.7	20
Band 8 - NIR	833.0	10
Band 8A - Vegetation red edge 4	864.0	20
Band 9 - Water vapor	943.2	60
Band 10 - Cirrus	1376.9	60
Band 11 - SWIR1	1610.4	20
Band 12 - SWIR2	2185.7	20

3.1.2. High spatial resolution imagery from the Google Earth dataset

Google Earth integrates spatial data from satellite images, aerial photographs, and geographic information. Currently, users can view high-resolution images of the Earth's surface on Google Earth (with a maximum size of 4800×2886 pixels), which are regularly updated over time. In other words, users can display images from the present or select a specific time in the past. This function makes Google Earth particularly useful for interpreting exposed coal sites, as its frequently updated imagery provides explicit visual representations of the ground surface.

3.2. Method

3.2.1. Processing procedure

The study follows the main steps outlined in Fig. 3 below: data collection, image preprocessing, spatial resolution enhancement, carbon index calculation and thresholding classification, noise filtering, accuracy assessment, result analysis, and validation.

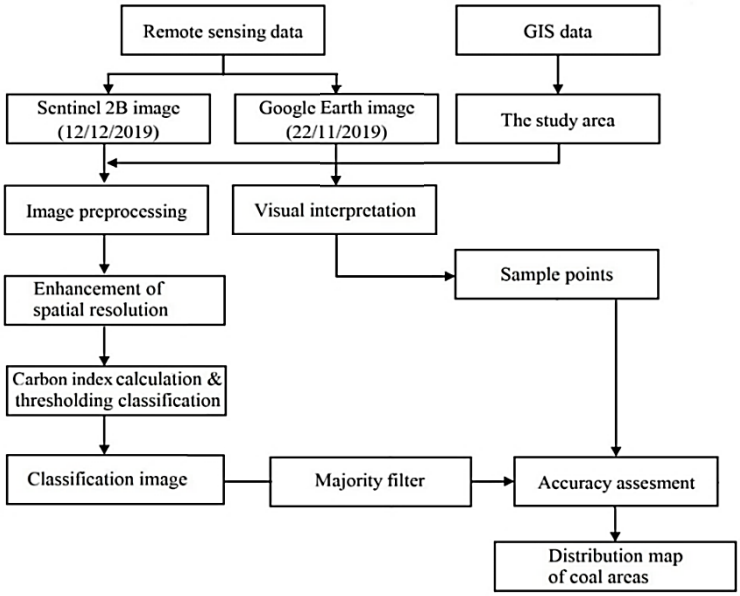


Figure 3. The processing procedure applied in this study

3.2.2. Pre-processing Sentinel-2 imagery

The Sentinel-2 image was atmospherically corrected for surface reflectance at level 2A and geometrically referenced to the UTM coordinate system (WGS84). It was then cropped to match the study area (Fig. 2) and used to calculate the spectral ratio for the acquisition date.

3.2.3. Spatial resolution enhancement

Each Sentinel-2 image consists of 13 spectral bands with varying spatial resolutions: 4 bands at 10 m, 6 bands at 20 m, and 3 bands at 60 m. The shortwave infrared (SWIR) bands have a spatial resolution of 20 m, while up to 4 bands in the visible and near-infrared (NIR) have a spatial resolution of 10 m. To enhance the level of detail, the spatial resolution of bands 11 and 12 is increased to 10 m using the Sen2Res algorithm while preserving the original reflectance values. The Sen2Res algorithm in SNAP software builds a model to analyze how information is shared between bands and what unique characteristics define each band. Next, the model is applied to unmix the 20 m and 60 m spatial resolution bands while preserving their reflectance values. This process enhances the ability to distinguish different entities more clearly (ESA, 2021b).

3.2.4. Carbon index and threshold identification

The carbon index is calculated as the ratio of reflectance values between two shortwave infrared bands (SWIR1 and SWIR2) based on the formula proposed by (Drury, 1993) for identifying the hydrothermal alterations in clay and aluminum minerals:

$$\varphi = (SWIR1-SWIR2)/(SWIR1+SWIR2) \quad (1)$$

(Mukherjee et al., 2018) the reflectance values of the SWIR1 band near the exposed coal sites are lower than those of the SWIR2 band due to the significant increase in mineral proportions on the surface. This causes the previously mentioned φ value to become negative. Each φ value (SWIR1, SWIR2) is treated as a function for identifying exposed coal sites using the segmentation technique as follows:

$F(\varphi) = 1$ if $\beta \leq \varphi\,(SWIR1, SWIR2) \leq \alpha$ (exposed coal sites)

$F(\varphi) = 0$ if $\beta > \varphi\,(SWIR1, SWIR2) > \alpha$ (other land cover types)

Here, α and β are the threshold values selected experimentally by analyzing the extent of exposed coal sites compared to other land cover types on a Sentinel-2 color composite image. This means that α and β values are determined based on the differences between the reflectance curve of coal and those of other land cover types, allowing exposed coal sites to be identified without relying on multiple indicators, additional methods, complex algorithms, or calculations.

Thus, the application of two shortwave infrared bands - corresponding to band 11 and band 12 of the Sentinel-2 image is incorporated into the following formula (2):

$$\varphi = (Band\ 11–Band\ 12)/(Band\ 11+Band\ 12) \quad (2)$$

Here, φ presents the carbon index. The spectral ratio φ (SWIR1, SWIR2) is an indicator for identifying exposed coal sites.

3.2.5. Visual interpretation

Figure 4 above presents the key elements of visual interpretation for remote sensing images. Based on these principles, coal extraction sites and coal yards are identified on the high-resolution image of the Google Earth dataset according to specific factors:

Site: The coal extraction sites are located near open-pit mines, where runoff water accumulates at the bottom of excavation holes. In contrast, coal yard sites are typically situated near coal extracting sites and are also found next to inland waterways ports/piers, conveyor belts, and railway stations, where raw coal loading and unloading activities occur (Fig. 5).

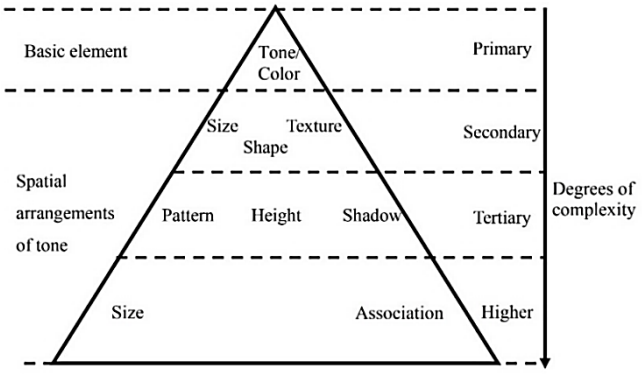


Figure 4. Elements for visual interpretation (Tiwari & Chatterjee, 2012)

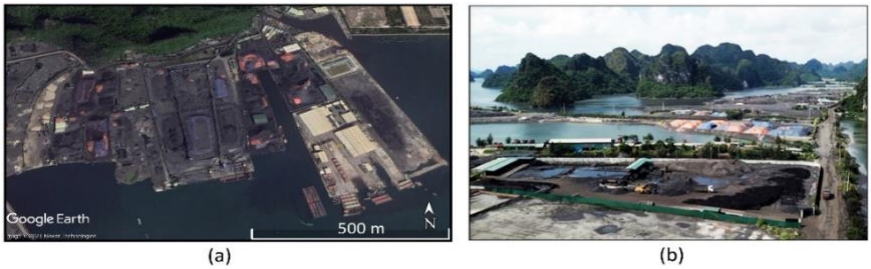


Figure 5. Coal yard sites on (a) the high-resolution image of the Google Earth dataset and (b) a photo from the real world

Structure and shape: Coal yard sites typically have a finer structure than waste rock dumps due to the coal screening process, which ensures proper selection and prevents mixing with other materials such as soil, rock, etc. Additionally, they often take on an inverted cone shape, as coal is gathered and piled up to conserve space and facilitate the loading and unloading (Fig. 6a). In contrast, coal extraction sites often exhibit a striped structure with undulating, convex grooves, resulting from the removal of rock layers to expose coal seams (Fig. 6b).

Figure 6. Features on the high-resolution image of the Google Earth dataset: (a) Coal yard sites, (b) Waste rock spoil area

Tone/color: Coal extraction sites or yard sites are typically black with a metallic sheen (Fig. 7a). This characteristic helps to distinguish coal from other features, such as vegetation and residential areas (Fig. 7b, c).

Association: It is an indirect indicator that recognizes objects or phenomena that cannot be directly perceived. However, their presence can be identified through the traces left on the ground. For example, coal extraction sites and

coal yards often contain various types of equipment, such as trucks, excavators, and conveyors, used for transporting coal from one location to another. Specifically:

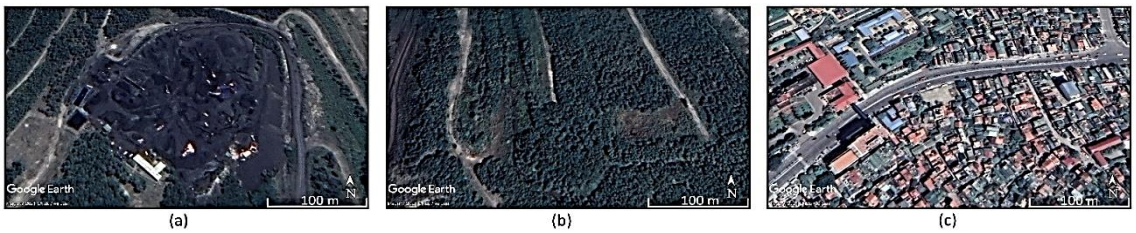


Figure 7. Features on the high-resolution image of the Google Earth dataset: (a) Coal extraction sites, (b) Vegetation, (c) Urban settlement

- Inland ports/piers are key transport hubs for processed coal products. Before being shipped to other locations, coal is loaded onto vessels, resulting in a high concentration of anchored vessels in the surrounding water area (Fig. 8a).

- Coal screening plants and thermal power plants consume large amounts of coal as raw materials for production. Coal conveyor belts and railways are commonly present inside these facilities to transport coal to production lines (Fig. 8b).

Figure 8. Features on the high-resolution image of the Google Earth dataset: (a) Vessels at coal ports (b) Conveyor belts in coal mines

3.2.6. Accuracy

To evaluate the accuracy of the established map, the thresholding classification results of coal index values are compared to those obtained through visual interpretation using overall accuracy and the Kappa coefficient. Specifically, overall accuracy is calculated as the ratio of the total number of correct sample points and the total number of sample points. The Kappa coefficient is determined using the following formula (Cohen, 1960):

$$K = [N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+} - x_{+i})]/[N^2 - \sum_{i=1}^{r}(x_{i+} - x_{+i})] \quad (3)$$

Here, r represents the number of rows and columns in the error matrix; x_{ii} is the number of pixels in row i and column i (on the main diagonal); x_{i+} is the total number of pixels in row i; x_{+i} is the total number of pixels in column i; and N is the total number of pixels in the image.

The Kappa coefficient typically ranges between 0 and 1, where $k \geq 0.75$ indicates high accuracy, $0.4 < k < 0.75$ indicates moderate accuracy, and $k \leq 0.4$ indicates low accuracy (Kirch, 2008).

4. Results

4.1. Exposed coal sites distribution

4.1.1. Carbon index value thresholding

The carbon index values obtained using formula (2) range from -1 to +1. To distinguish exposed coal sites from other features, these values were segmented based on their corresponding entities in a color composite image (Red-Green-Blue) created using shortwave infrared and near-infrared bands (b12-b11-b5) from the Sentinel-2 image. Experiments show that the threshold values of α and β range from -0.19 to -0.06, effectively representing the extent of exposed coal sites. The carbon index accurately distinguishes coal from other features (Fig. 9). Notably, these thresholds enable the clear separation of exposed coal sites from urban areas, avoiding the confusion observed in the study (Raju, 2016), where visible spectrum bands were used.

4.1.2. Generalization of isolated pixels

After thresholding classification, small and discrete clusters of pixels representing exposed coal sites are commonly observed in the results. This noise can impact the reader's ability to interpret the information accurately. Several methods have been developed to address this issue. The first approach improves classification accuracy by incorporating the spectral-spatial properties of entities (Bruzzone & Carlin, 2006; Huang & Zhang, 2013), while the second approach applies post-classification processing (Lv, Zhang, & Benediktsson, 2017; Su, 2016; Tu, Aa, Gemeren, & Veltkamp, 2013).

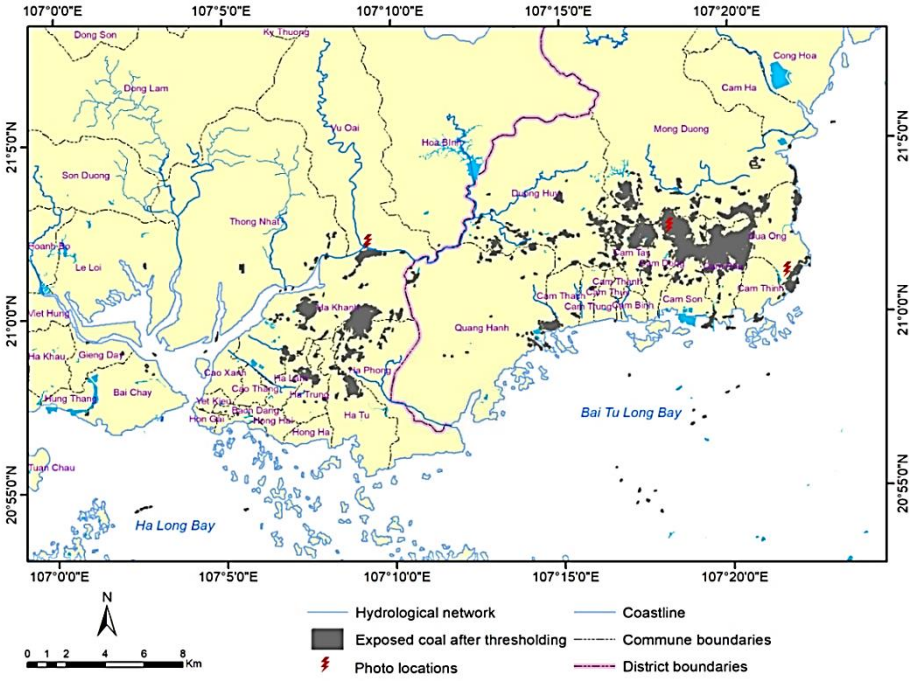


Figure 9. Carbon index thresholding classification

Studies indicate that the sliding window technique is usually used to analyze neighboring pixels when reassigning values to the central pixel. This method effectively removes isolated pixels while preserving classification accuracy. The present study adopts the second approach, applying a Majority filter with a 5 × 5 dimension to the segmented classification image due to its effectiveness in comparison to other

dimensions, which removes isolated pixels while preserving the shape of features. The resulting image highlights exposed coal sites while minimizing noise (Fig. 10).

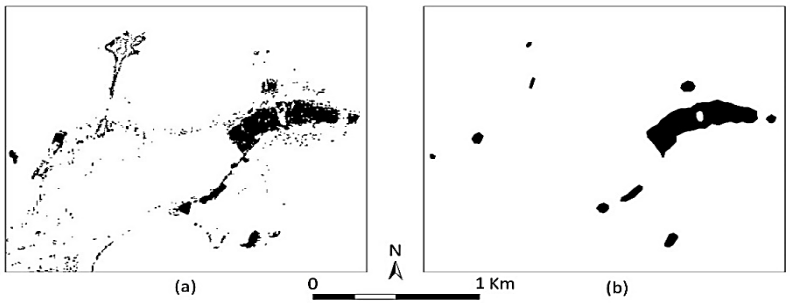


Figure 10. Coal yard sites on the Dien Vong riverbank: (a) before and (b) after filtering

4.1.3. Exposed coal site distribution

Figure 11 illustrates the distribution of exposed coal sites after applying the Majority filter. Specifically, the exposed coal sites are divided into two main clusters located at the center of the study area, approximately 3 km from the coastline. These exposed coal sites correspond to opencast coal extraction sites and coal yards. In Ha Long City, exposed coal sites are concentrated in several wards, including Ha Khanh, Ha Phong, Ha Tu, Ha Trung, Ha Lam, and Thong Nhat commune. In Cam Pha City, they are primarily found in Quang Hanh, Cam Thach, Cam Trung, Cam Tay, Cam Dong, Cam Son, Cam Phu, Cam Thinh, Cua Ong, Mong Duong, and Duong Huy commune. Notably, exposed coal sites are more widespread in Cam Pha city.

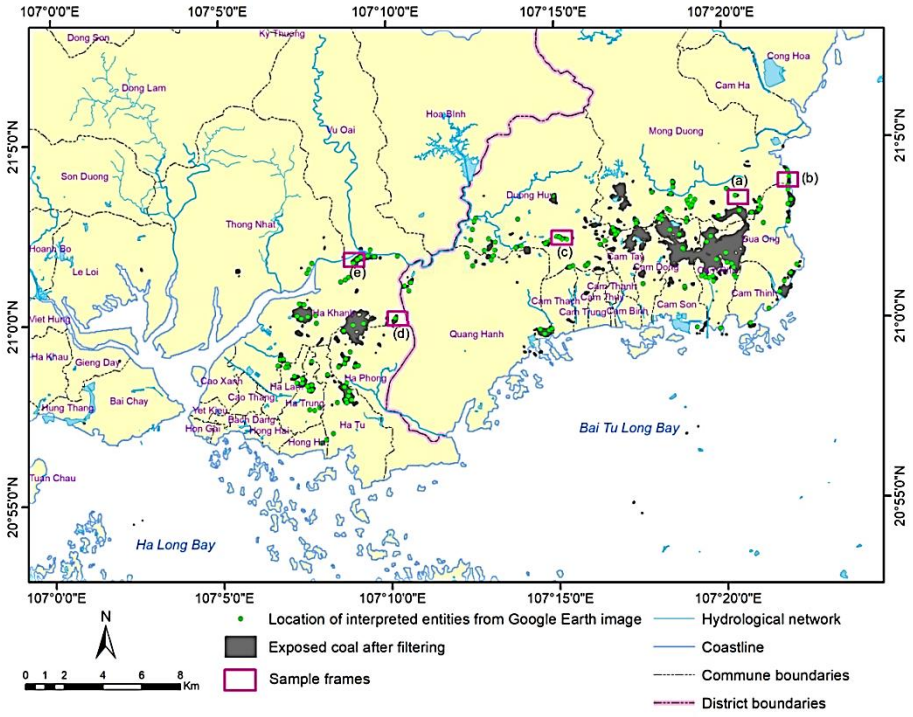


Figure 11. Exposed coal site distribution in the Ha Long - Cam Pha region. The pink boxes indicate the locations of sample frames shown in Fig. 12 a-e

This pattern aligns with the development plan of the Vietnam National Coal and Mineral Industries Group, as Ha Long City is gradually restricting opencast coal mining operations in preparation for their closure in the coming years (NIURP, 2012). In both cities, large, exposed coal sites are typically associated with coal extraction sites. In contrast, smaller sites are found near inland waterway ports, piers, coal screening facilities, and thermal power plants.

4.2. Validation

4.2.1. Field trip

Two field trips were conducted in September 2020 to recognize specific land cover types in the study area and compare the results with real-world conditions. This serves as a basis for interpreting high-resolution images of the Google Earth dataset. Photos were taken at several coal extraction and yard sites (Fig. 12), where their locations align with the exposed coal sites identified in the segmented carbon index image.

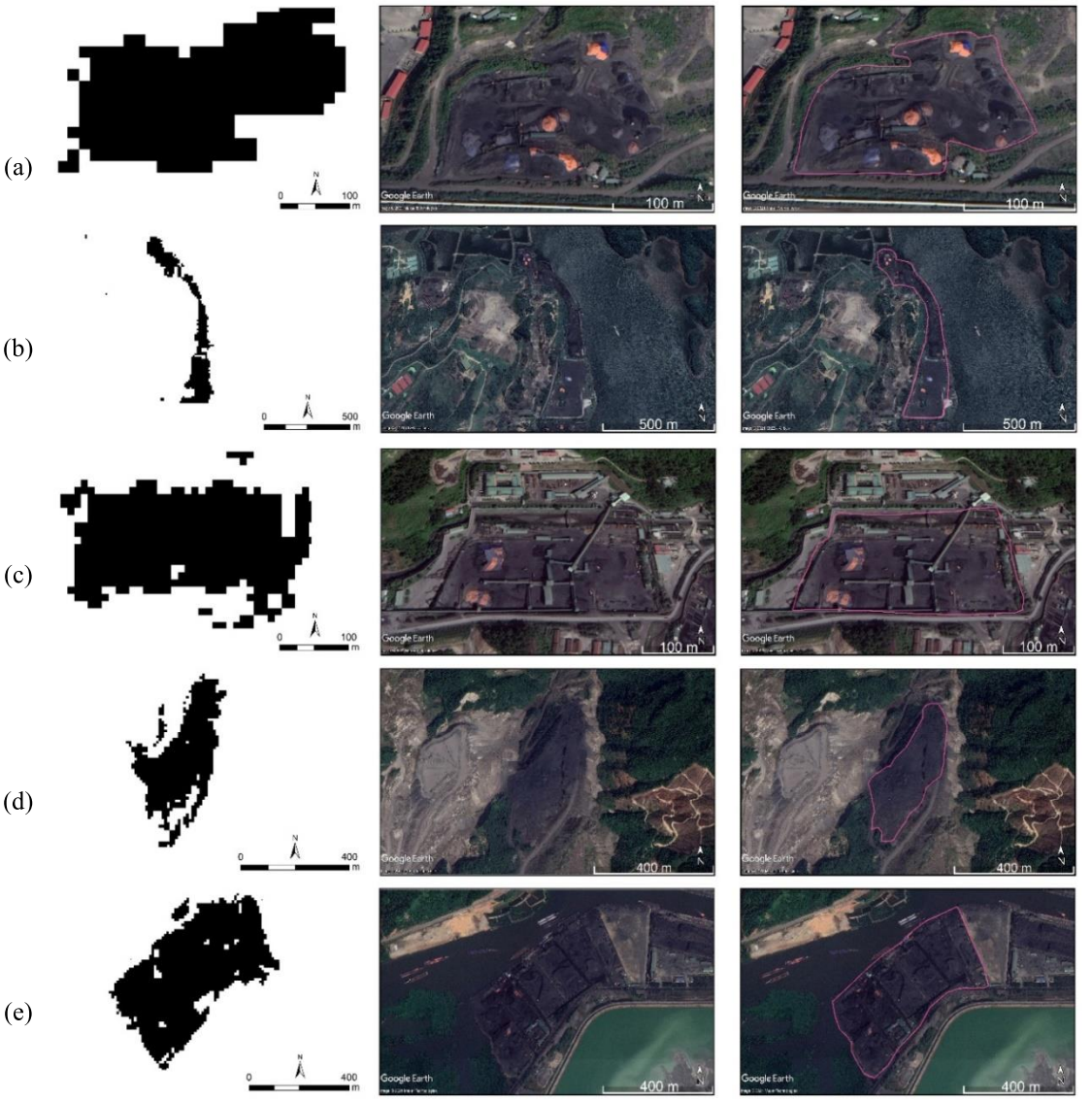
4.2.2. Comparison with high-resolution images of the Google Earth dataset

The validation is based on comparing carbon index thresholding classification using the Sentinel-2 image (captured on December 12, 2019) and visual interpretation results from a high spatial resolution image (captured on November 22, 2019) of the Google Earth dataset. A set of 245 random samples was extracted from the Google Earth dataset, of which 161 samples corresponded to exposed coal sites, while 84 samples represented other land cover types. These 245 samples were then superimposed onto the carbon index thresholding classification image to assess the positional alignment between the two data layers. The sample locations are presented in Fig. 11, and the statistical results are shown in Table 3. The statistical analysis indicates that the overall accuracy reached 87.35%, while the Kappa coefficient was 0.73. To illustrate the validation framework, five specific samples are highlighted in Fig. 13 rather than displaying the entire set of 245 samples.

Figure 12. (a) The coal yard site at Lang Khanh Port is located on the left bank of the Dien Vong River. (b) Coal extraction site at Deo Nai coal mine. (c) The coal yard site at Cua Ong coal screening plant and Cam Pha thermal power plant are located onshore in Bai Tu Long Bay. For the photo locations, refer to Fig. 9

Table 3. Matrix used for accuracy evaluation

		Google Earth dataset					
		Other entities		Exposed coal sites		Total	
		samples	%	samples	%	samples	%
Thresholding classification of carbon index	Other entities	75	30.61	22	8.78	97	39.59
	Exposed coal sites	9	3.68	139	56.73	148	60.41
	Total	84	34.29	161	65.71	245	100



Exposed coal sites identified using carbon index thresholding classification

The natural composition of the Google Earth dataset

Exposed coal sites were identified by visual interpretation with the Google Earth dataset in the background

Figure 13. Comparison of the spatial extent of exposed coal sites identified using two different methods. For the location of frame, refer to Fig. 11

5. Discussions

Several limitations exist for the proposed processing approach for optical imagery. Some exposed coal patches are missing from the coal distribution map, or their dimensions appear smaller than the corresponding entities in the high-resolution image of the Google Earth dataset. This can be attributed to two main reasons: -/ spatial resolution differences: the Sentinel-2 image has a lower spatial resolution compared to those of the Google Earth dataset, making it more challenging to identify smaller entities; -/effect of post-processing: after thresholding classification, the Majority filter reduced the size of objects or in some cases, completely removed them. As a result, the validation of Sentinel-2 image against visual interpretation using the Google Earth dataset yielded an overall accuracy of 87.35%.

All stages of coal mining activities, including stripping, extraction, dumping, screening, and transportation, have significant environmental impacts, one of the most concerning being water pollution. Although wastewater originates from various sources, including groundwater from opencast and underground coal mines, as well as effluents from screening plants, it is collected for treatment. However, only 74% of this wastewater is treated before being discharged into the environment (KOEI, 2014). During the rainy season, surface runoff from coalfields and waste rock dump areas still directly enters the drainage network. As a result, coal slurry is carried into the bay, contributing to elevated total suspended solids (TSS) levels, which often exceed the allowable limits set by Vietnamese regulations for the protection of aquatic life (Dao, Bui, Nguyen, Nguyen, & Bui, 2016). A study conducted by Tran et al. (2012) found that each year, Ha Long and Bai Tu Long bays receive large amounts of pollutants from industry, livestock, domestic use, and coal mining.

For example, in the Cam Pha area, aside from domestic waste, livestock contributes 13.9 thousand tons of Chemical Oxygen Demand (COD) and 1.5 thousand tons of Biochemical Oxygen Demand (BOD_5). Additionally, coal mining also releases 70 thousand tons of Total Suspended Solids (TSS) in the form of slurry. At the Cua Luc estuary, slurry deposition ranges from 0.1 to 3.5%, with some areas reaching 10%. Figure 14 illustrates the dispersion of coal slurry in seawater around coal ports/piers, screening plants, and estuaries, which are connected to coalfields through designated discharge channels for mine water. This highlights the close relationship between exposed coal sites and the concentration of coal slurry in the seawater of Bai Tu Long and Ha Long bays.

Illegal coal mining remains a persistent issue in the remote areas of Ha Long and Cam Pha cities, yet it is challenging to monitor using traditional investigative methods. This results in adverse environmental impacts and the depletion of natural resources. However, illegal coal mining is often associated with coal yard sites, making it possible to identify their locations by comparing them with the exposed coal site distribution map generated through carbon index thresholding classification. By combining this approach with the official registration data of the Quang Ninh Department of Natural Resources and Environment, authorities can assess the legal status of coal extraction sites and identify individuals and institutions involved in violations.

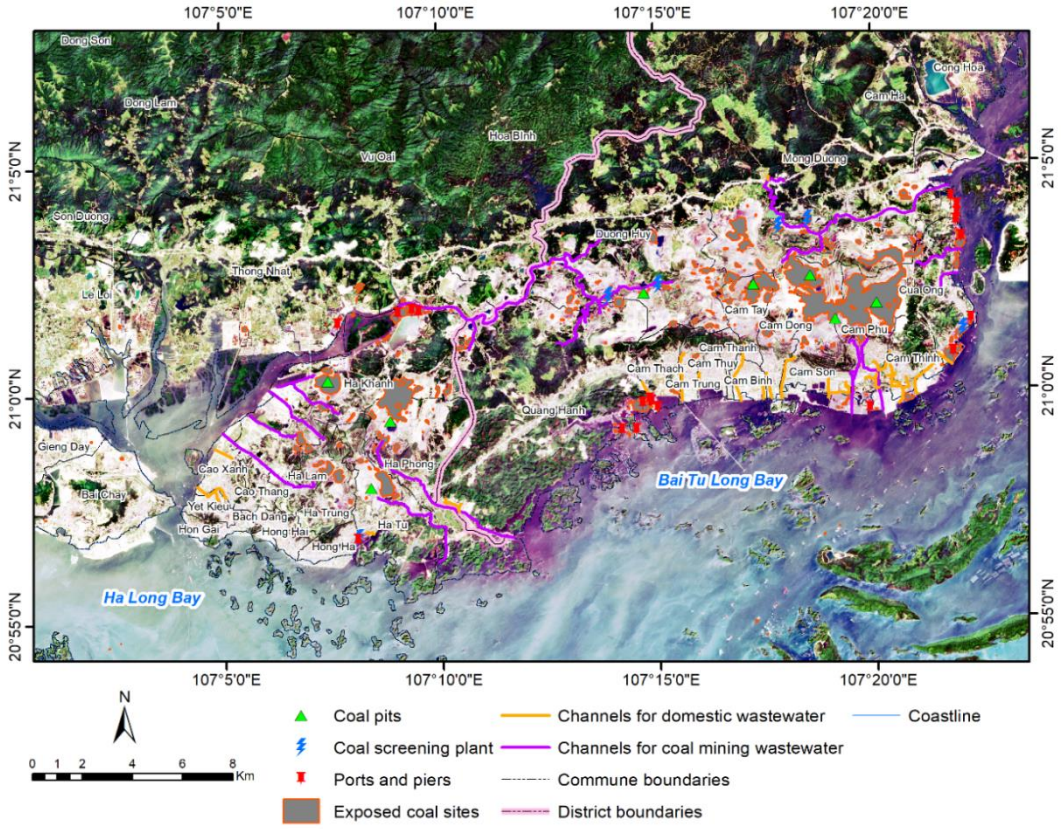


Figure 14. Coal slurry in seawater (dark violet) in the Ha Long and Cam Pha coastal regions. The natural color composite (b4-b3-b2) of the Sentinel-2 image, captured on December 23, 2019, is shown in the background

6. Conclusions

The study has demonstrated the applicability of Sentinel-2 data in identifying exposed coal sites. The calculation of the carbon index from shortwave infrared bands, combined with segmentation within the range of [-0.19; -0.06], enables automated mapping of the spatial distribution of opencast coal extraction sites and coal yard sites in the coastal area of Ha Long and Cam Pha cities (Quang Ninh province). The validation achieved a satisfactory accuracy of 87.35%, based on visual interpretation of the Google Earth dataset captured during the same period. This enables the estimation of the area of exposed coal sites and the assessment of the legal status of coal extraction sites, among other tasks, aiding authorities in implementing effective plans and policies for coal mining management in this territory. Enhancing the spatial resolution of shortwave infrared bands to 10 m improves the detection accuracy of exposed coal sites; however, this resolution remains insufficient, as small exposed coal sites may still be too minor to be distinguished in Sentinel-2 images. Therefore, future research should explore advanced methods to achieve higher accuracy.

Acknowledgments

The authors acknowledge Hanoi National University of Education and Ho Chi Minh City University of Education for their support in this study. We also thank the European

Space Agency for providing Sentinel images through the Copernicus data hub.

Author declaration

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

References

Azeez A.-S.H.A., Mukhitdinov S., 2020. Land use land cover change detection in the mining areas of V.D. Yalevsky coal mine-Russia. E3S Web Conference, 192, 04021. https://doi.org/10.1051/e3sconf/202019204021.

Bruzzone L., Carlin L., 2006. A Multilevel Context-Based System for Classification of Very High Spatial Resolution Images. IEEE Transactions on Geoscience and Remote Sensing, 44(9), 2587−2600. https://doi.org/10.1109/TGRS.2006.875360.

Cao T.T.T., Do C.T., Le V.N., Pham T.K., Nguyen V.B., Dinh H.N., 2020. Assessment of Sea Water Quality in some Limestone Island and Archipelagos Areas, Vietnam. VNU Journal of Science: Earth and Environmental Sciences, 36(1), 70−78. https://doi.org/10.25073/2588-1094/vnuees.4556.

Chen W., Li X., Wang L., 2020. Fine Land Cover Classification in an Open Pit Mining Area Using Optimized Support Vector Machine and WorldView-3 Imagery. Remote Sensing, 12(1), 82. https://doi.org/10.3390/rs12010082.

Cira D., Dastur A., Jewell H., Kilroy A., Lozano N., Phan H.T.P., Wang H.G., 2011. Vietnam Urbanization Review: Technical Assistance Report. Retrieved from Washington DC: http://documents.worldbank.org/curated/en/225041468177548577.

Cohen J., 1960. A Coefficient of Agreement for Nominal Scales. New York University.

Dao B.T., Bui D.C., Nguyen T.T.H., Nguyen T.P.T., Bui T.N., 2016. Modeling spatial distribution of total suspended solids concentration in Ha Long Bay water during the first quarter of 2016 using co-kriging interpolation and auxiliary data from Landsat 8 imagery. Paper presented at the The 7th International Symposium Hanoi Geoengineering Hanoi, 153−156.

Demirel N., Emil M.K., Duzgun H.S., 2010. Surface coal mine area monitoring using multi-temporal high-resolution satellite imagery. International Journal of Coal Geology, 86(1), 3−11. https://doi.org/10.1016/j.coal.2010.11.010.

Drury S.A., 1993. Image Interpretation in Geology. London, UK: Chapman & Hall, p.283.

ESA, 2021a. MultiSpectral Instrument (MSI) Overview. Retrieved from https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument.

ESA, 2021b. Sen2Res. Retrieved from https://step.esa.int/main/snap-supported-plugins/sen2res/.

Gao Y., Kerle N., Mas J.-F., 2009. Object-based image analysis for coal fire-related land cover mapping in coal mining areas. Geocarto International, 24(1), 25−36. https://doi.org/10.1080/10106040802395648.

Giri C.P., 2012. Remote Sensing of Land Use and Land Cover Principles and Applications. New York: CRC Press, pp.480.

Hens L., et al., 2000. Land Cover Changes in the Extended HA Long City Area, Northeastern Vietnam During the Period 1988–1998. Environment, Development and Sustainability, 2, 235–252. https://doi.org/10.1023/A:1011466108499.

Huang X., Zhang L., 2013. An SVM Ensemble Approach Combining Spectral, Structural, and Semantic Features for the Classification of High-Resolution Remotely Sensed Imagery. IEEE Transactions on Geoscience and Remote Sensing, 51(1), 257−272. https://doi.org/10.1109/TGRS.2012.2202912.

Huynh S., 2002. Hydrodynamic study of Ha Long Bay. (Bachelor of Engineering), University of Western Australia, Perth, Australia, p.81.

Kirch W., 2008. Kappa Coefficient. In Encyclopedia of Public Health. Dordrecht: Springer, p.1601.

KOEI N., 2014. Environmental planification of Quang Ninh province to 2020 and vision to 2030. Retrieved from Ha Long: https://www.quangninh.gov.vn/So/sotainguyenmt/Lists/TinTuc/Attachments/598/.

Lv Z., Zhang X., Benediktsson J.A., 2017. Developing a general post-classification framework for land-cover mapping improvement using high-spatial-resolution remote sensing imagery. Remote Sensing Letters, 8(7), 607−616. https://doi.org/10.1080/2150704X.2017.1306137.

Mai A., Quyen D., Luong D.T., 1983. General basics on geology and mineral of Quang Ninh coal basin. Vietnam Geologic Journal, 10, 7−13.

Mao Y., Ma B., Liu S., Wu L., Zhang X., Yu M., 2014. Study and Validation of a Remote Sensing Model for Coal Extraction Based on Reflectance Spectrum Features. Canadian Journal of Remote Sensing, 40(5), 327−335. https://doi.org/10.1080/07038992.2014.979486.

Mukherjee J., Mukherjee J., Chakravarty D., Aikat S., 2018. A Novel Index to Detect Opencast Coal Mine Areas From Landsat 8 OLI/TIRS. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(3), 891−897. https://doi.org/10.1109/JSTARS.2019.2896842.

Mustafin M.G., Tran T.S., Tran M.H., 2019. Comprehensive impact assessment development of the Coal field Campha in Vietnam to the coastal territory. In: Construction and Architecture: Theory and Practice of Innovative Development, (Kislovodsk, Russian Federation: IPO Publishing), 1−7.

Nguyen V.T., 2015. Research on topographic changes in relationship with coastal ecosystems in Quang Ninh province on the basis of application of remote sensing and GIS technology. In, Faculty of Geography. Hanoi: VNU University of Science, p.162.

NIURP, 2012. City development strategies for medium size cities in Vietnam: Can Tho and Ha Long. In Technical report, Hanoi: National Institute of Urban and Rural Planning, p.205.

Pandey M., Mishra A., Swamy S.L., Anderson J.T., Thakur T.K., 2025. Machine learning-based monitoring of land cover and reclamation plantations on coal-mined landscape using Sentinel 2 data. Environmental and Sustainability Indicators, 25, 100585. https://doi.org/10.1016/j.indic.2025.100585.

Pham A.V.T., Do D.M., Pham H.T.T., Phan T.T., Pham H.N., 2020. Application of relative air pollution index (RAPI) a new method for aggregate assessment of current air pollution in Cam Pha coal mining area, Quang Ninh province, Vietnam. Environmental Monitoring and Assessment, 192(411). https://doi.org/10.1007/s10661-020-8242-1.

Raju A., 2016. Remote sensing based coal fire studies in Jharia coalfield, India. In, Earth Science. Roorkee, Uttarakhand: Indian Institute of Technology Roorkee, p.189.

Song W., Song W., Gu H., Li F., 2020. Progress in the Remote Sensing Monitoring of the Ecological Environment in Mining Areas. International Journal of Environmental Research and Public Health, 17(1846), 1−17. https://doi.org/10.3390/ijerph17061846.

Su T.-C., 2016. A filter-based post-processing technique for improving homogeneity of pixel-wise classification data. European Journal of Remote Sensing, 49(1), 531−552. https://doi.org/10.5721/EuJRS20164928.

Tiwari K.N., Chatterjee C., 2012. Remote Sensing and GIS Application. Retrieved from http://ecoursesonline.iasri.res.in/course/view.php?id=53.

Tong D.T., et al., 2005. Lexicon of geological units of Vietnam. Hanoi: Hanoi National University Press, p.520.

Tran, D.T., et al., Environmental capacity of Ha Long - Bai Tu Long Bay, 2012. Hanoi: Natural Sciences and Technology Publishing House, p.279.

Tran D.T., 1998. The geological history of Ha Long Bay. Hanoi: The World Publishing House, p.94.

Tu Z., Aa N.d., Gemeren C., Veltkamp R.C., 2013. A combined post-filtering method to improve accuracy of variational optical flow estimation. Pattern Recognition, 47(5), 1926−1940. https://doi.org/10.1016/j.patcog.2013.11.026.

Van D., 2015. Illegal coal mining continues. Vietnam News. Retrieved from https://vietnamnews.vn/society/273179/illegal-coal-mining-continues.html.

Waltham T., 1998. Limestone Karst of Ha Long Bay Vietnam: An Assessment of the Karst

Geomorphology of the World Heritage Site for The World Conservation Union and The Management Department of Ha Long Bay. Nottingham: Nottingham Trent University, p.82.

Xiao D., Le B.T., Mao Y., Jiang J., Song L., Liu S., 2016. Research on Coal Exploration Technology Based on Satellite Remote Sensing. Journal of Sensors, 9. https://doi.org/10.1155/2016/8214801.

Yang E., Ge S., Wang S., 2018. Characterization and Identification of Coal and Carbonaceous Shale Using Visible and Near-Infrared Reflectance Spectroscopy. Journal of Spectroscopy, 13. https://doi.org/10.1155/2018/2754908.

Yang Z., He T., Zhang J., Zhao Y., 2024. A novel index for exposed coal mapping using Landsat imagery. Ecological Indicators, 166, 112395. https://doi.org/10.1016/j.ecolind.2024.112395.

Yuan X., Xing L., Gao W., Gao P., Zhao X., Chang J., 2019. Object-oriented monitoring of coal mine land use change. IOP Conference Series: Materials Science and Engineering, 592(1), 012187. https://doi.org/10.1088/1757-899X/592/1/012187.