



Coastal vulnerability assessment using the machine learning tree-based algorithms modeling in the north coast of Java, Indonesia

Fajar Yulianto¹ · Mardi Wibowo¹ · Ardila Yananto¹ · Dhedy Husada Fadjar Perdana¹ · Edwin Adi Wiguna¹ · Yudhi Prabowo¹ · Nurkhalis Rahili¹ · Amalia Nurwijayanti¹ · Marindah Yulia Iswari¹ · Esti Ratnasari¹ · Amien Rusdiutomo¹ · Sapo Nugroho¹ · Andan Sigit Purwoko¹ · Hilmi Aziz¹ · Imam Fachrudin¹

Received: 30 May 2023 / Accepted: 22 October 2023 / Published online: 9 November 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

The north coast of Java is the center of economic activity in Indonesia. This area is dynamic and sensitive to various geo-bio-physical aspects. Therefore, a vulnerability study in this area is necessary. This study proposes a machine learning tree-based algorithms modeling approach for Coastal Vulnerability Assessment (CVA) and mapping. The tree-based algorithms used are Gradient Tree Boost (GTB), Classification and Regression Trees (CART), and Random Forest (RF). The study utilized the Google Earth Engine (GEE) platform and twelve variables as input. The prediction results of each of these modeling algorithms have been compared and evaluated to determine the most optimal performance and accuracy. Reference data was obtained from the Ministry of Maritime Affairs and Fisheries of the Republic of Indonesia (KKP). Approximately 70% of the reference data was allocated for training, while the remaining 30% was designated for validation. The CVA assessment yielded overall accuracies of 80.22%, 77.40%, and 71.18% based on the RF, GTB, and CART algorithms, respectively. Meanwhile, the Kappa Index for these three algorithms was 0.72, 0.67, and 0.58, indicating that the models have adequately classified the data. The research outcomes are anticipated to offer insights into the potential utilization of machine learning technology for vulnerability assessment and mapping, contributing to the management of coastal environmental issues.

Keywords Geospatial data · Machine learning · Tree-based algorithms · Vulnerability · North coast of Java · Indonesia

Introduction

A coastal area is a natural landscape categorized based on the maritime boundaries. It is indicated by a breaking zone inland on the coastal alluvial plain landscape. These areas also have abundant natural resources and play an important role in sustainable development influenced by various factors (Rizzo et al. 2020; Nurshodikin and Saputra 2021). These factors are ecological production, population, industrial compatibility, transportation, waste disposal, military strategy development, and others (Noor and Abdul Maulud 2022). In addition, the coastal area changes are extremely

dynamic with various physical processes such as coastal erosion, sedimentation, increasing sea level, land subsidence, inundation of wetlands and lowlands, increased flooding, coastal destruction, salinity condition of estuaries and aquifers (Marfai and King 2008; Gopinath et al. 2016; Yankey et al. 2020; Arjasakusuma et al. 2021; Jafariroodsari et al. 2021). Thus, numerous changes in the coastal area cause coastal areas to be vulnerable to coastal hazards, damage, and environmental pollution (Atikawati et al. 2019; Kathiravan et al. 2019).

Indonesia belongs to a tropical environment, and it is the largest archipelagic state on Earth, consisting of more than 17,500 islands with a shoreline extending over 81,000 km (Nikijuluw 2017; Sui et al. 2020; World Bank and Asian Development Bank 2021). The tropical environments exhibit distinct processes compared to polar and temperate regions, which can influence coastal conditions. Furthermore, a coastal area in tropical environments possesses unique characteristics, including high temperatures, rainfall, solar radiation, and runoff. In addition, these processes are influenced

Communicated by: H. Babaie

✉ Mardi Wibowo
mardi.wibowo@brin.go.id

¹ Research Center for Hydrodynamics Technology, National Research and Innovation Agency (BRIN), Jl. Hidro Dinamika, Sukolilo, Surabaya, East Java 60112, Indonesia

by water discharge, sedimentation, and solutes from tropical rivers (Nittrouer et al. 1995; Val et al. 2006). The other study highlights that tropical coastal ecosystems are characterized by mangrove forests, seagrass beds, coral reefs, and high marine biodiversity and productivity (Done and Jones 2006).

The coastal areas in tropical environments are vulnerable to climate dynamics and sea level rise (Sui et al. 2020; World Bank and Asian Development Bank 2021). Even storms, rising sea levels, and land conversion can impact and reduce the resilience of the region's natural and social systems (Pramova et al. 2013). It can affect the potential commercial deficit for the coastal sector, amounting to IDR 6.7 trillion and approximately IDR 400.8 trillion for the marine sector (Kementerian PPN/Bappenas 2020). Furthermore, over 60% of Indonesia's residents live in coastal zones, contributing over 24% to the national economy regarding marine-based activities (Purwaka and Sunoto 1997).

Several variables affect coastal vulnerability, including geological, physical, and sociocultural factors (Özyurt and

Ergin 2009). The geological variable assesses the potential for erodibility, encompassing shoreline change, geology, and geomorphology (Samanta and Paul 2016; Rizzo et al. 2020; Natarajan et al. 2021). The physical variable indicates the possibility of coastal flooding, comprising mean wave height, increasing sea level, mean tidal range, and significant wave height (Pramanik et al. 2016). Meanwhile, the variables represent sociocultural characteristics that influence coastal environmental conditions, such as economic activity, infrastructure, residents, road path, land use, and more. (Bukvic et al. 2020; Noor and Abdul Maulud 2022; Tanim et al. 2022). Coastal management is critical in estimating coastal environmental and socio-economic factors dynamics, providing essential regional and local services (Astsatryan et al. 2022).

The study area is located in the north coast of Java, Indonesia (Fig. 1). This location holds strategic importance for economic activities. However, it is also highly vulnerable to increasing sea levels, land subsidence, and groundwater

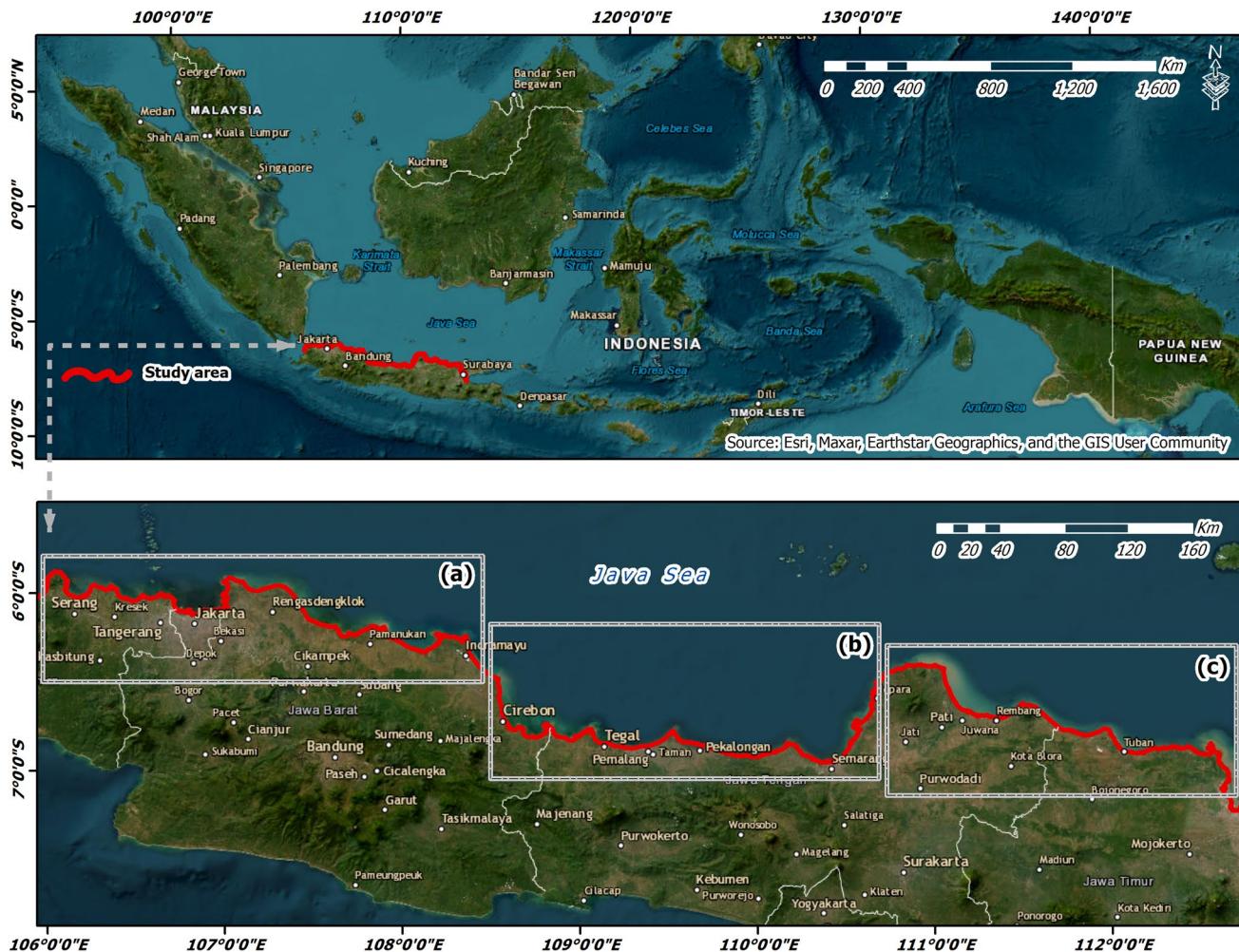


Fig. 1 The study area is located in the north coast of Java, Indonesia. **a** the study area in the western part. **b** the study area in the middle part. **c** the study area in the eastern part

abstraction, posing a threat to the sustainability of cities in these coastal areas (Wu 2021). Most coastal areas, including the north coast of Java, have abundant water resources (both surface water and groundwater) and gentle terrain, making them suitable for urban development, industrial estates, office centers, ports, tourism, fisheries, agriculture, and settlements (Ridwansyah et al. 2020). Consequently, large-scale land conversion and uncontrolled natural resource extraction contribute to environmental problems, such as pollution, coastal erosion, tidal waves, land subsidence, inundation, and social conflicts (Firman 2009). Global warming further exacerbates these challenges, increasing the risk of tidal flooding (Solihuddin et al. 2021; Mandal et al. 2021).

Many studies have been conducted on the coastal environment of the north coast of Java. These studies have focused on various aspects, including coastal erosion (e.g., Marfai 2011a; Marfai 2011b; Achiari et al. 2015; Suwandana 2019; Suhardi et al. 2020; Warnadi et al. 2020; Solihuddin et al. 2021), coastal inundation (e.g., Wibowo et al. 2015), shoreline changes (e.g., Prasetyo et al. 2019; Khadijah et al. 2020; Radiarta et al. 2022) and land subsidence (e.g., Abidin et al. 2013; Yuwono et al. 2016; Sidiq et al. 2021).

Coastal vulnerability can be assessed and mapped using the CVA method. This map is a valuable tool for determining coastal protection priorities and assisting decision-makers in shaping coastal management policies. (Nguyen 2015; Hossain et al. 2022). Therefore, CVA is crucial in addressing environmental challenges, protecting coastal ecosystems, and mitigating coastal disasters (Aliño et al. 2013; Krishnan et al. 2019; Mahamoud et al. 2022). Noor and Abdul Maulud (2022) conducted a comprehensive literature review of CVA methods. According to their findings, CVA methods can be categorized into four types: indexes, indicators, GIS, and dynamic computer models.

The index-based approach method has been employed in various studies, including those by Gornitz (1990), Sankari et al. (2015), Barman et al. (2016), Ghazali et al. (2018), De Serio et al. (2018), El-Shahat et al. (2021), Luu et al. (2021). The indicator-based method has been utilized in studies conducted by Huu and Huynh (2018) and Rizzo et al. (2020). The GIS-based method has been employed with the use of technology in research by Beluru Jana and Hegde (2016), Emran (2016), Malik and Abdalla (2016), Ghoussein et al. (2018), Hammid et al. (2018), Chakraborty (2021), Nath et al. (2021). Dynamic computer models have been used in research by Bagheri et al. (2021), Liang et al. (2021), Tran et al. (2022), Ndehedehe et al. (2022), Shaikh et al. (2022). These descriptions illustrate that these four methods have been widely applied in various locations. In our research, we employ machine learning with a tree-based algorithm to assess coastal vulnerability on the northern coast of Java.

Studies on CVA are primarily conducted in cities on Java, Indonesia's north coast. Several of these studies have

employed various methods, including scoring, weighting, and the Analytic Hierarchy Process (AHP), in conjunction with Geographic Information System (GIS) analysis. For example, the vulnerability of the north coast of Central Java was analyzed using the Coastal Risk Assessment Framework (CRAF) method, as assessed by Prasetya (2021). In a separate study, Kasim et al. (2011) investigated the vulnerability of the north coast of Indramayu, West Java, employing an integrated approach that combined the Coastal Vulnerability Index (CVI), Multi-Criteria Analysis (MCA), and GIS techniques. Additionally, Bengen and Tahir (2012) mapped coastal vulnerability to climate change in five major cities on the north coast of Java. Handiani et al. (2022) assessed and compared Subang Regency, West Java coastal vulnerability, using the CVI and CVI weighted (CVIw) methods. Furthermore, Handiani et al. (2022) utilized the CVIw method to study coastal vulnerability on the north coast of Java. Lastly, Anwar et al. (2020) studied coastal vulnerability along the north and south coasts of West Java based on oceanographic and ecosystem parameters.

The vulnerability of the national coast and the five major cities on the north coast of Java has been studied and mapped using the CVI method by the Ministry of Maritime Affairs and Fisheries, Republic of Indonesia (2009). The results of these studies have shown that the vulnerability on the north coast of Java is classified as high and very high. This study employed indicators and index-based methods. Several positive and negative consequences of indicators and index-based techniques have been identified in previous studies. One of the advantages is the simplicity of implementing these indicators and index-based methods (Noor and Abdul Maulud 2022). Index-based techniques provide a straightforward way to classify different options during decision-making processes (Giannakidou et al. 2020). They have gained wide acceptance as the preferred method for vulnerability (Blasiak et al. 2017; Debortoli et al. 2019). On the other hand, a weakness of existing methods is that they require significant time and effort to add or update new parameters. Despite the rapid changes in many parameters and the frequent emergence of new ones, CVA must be updated promptly.

Now, advances in geospatial technology have been developed and widely used across various sectors. Geospatial Artificial Intelligence (Geo-AI) represents one of these advancements in geospatial technology, integrating geospatial data, GIS, and AI (Kaur and Sood 2020). The Geo-AI approach can serve as an alternative to address the limitations and challenges encountered when conducting vulnerability assessments in the field.

According to McLaughlin et al. (2010), the CVA variables were categorized into three groups: geologic vulnerability (coastal feature), physical vulnerability (coastal processes), and socio-economic vulnerability. This study

primarily focuses on the geo-bio-physical aspects of coastal vulnerability variables in the study area. Generally, previous studies utilized the seven variables found in the CVI-Gornitz model (Gornitz 1990). These variables are geomorphology, coastal slope, sea level change, shoreline change, mean tide range, and mean wave height. Meanwhile, this research used 12 parameters by adding other parameters considered locally to influence coastal vulnerability. These additional parameters are coastal elevation, bathymetry, geological land subsidence, land use, and flood inundation. Moreover, most of these parameters are derived from multi-source geospatial data that is freely accessible to the public. By considering as many parameters as possible, coastal vulnerability studies and mapping results are better and more accurate.

Currently, most studies and mapping of coastal vulnerability use indexes, indicators, GIS, and dynamic computer models (Noor and Abdul Maulud 2022). This study proposes a Geo-AI approach that employs machine learning methods based on tree-based algorithms for vulnerability assessment and mapping on the north coast of Java, Indonesia. This study utilized the Google Earth Engine (GEE) cloud computing platform. Using GEE, various geospatial data sources can be obtained and integrated with classic machine learning algorithms, including tree-based algorithms. This approach allows for faster and easier vulnerability assessment and mapping, especially when updates or additional parameters

are required. This study represents the first attempt to integrate multi-source geospatial data for CVA and mapping using a tree-based algorithms machine learning approach. Previous methods for determining CVA and mapping relied on index-based approaches, indicators, GIS, and dynamic computer models.

The findings of this study provide insight and prospects for applying machine learning methods and improving CVA and mapping techniques. Another benefit of this study is that the CVA output in the research area can be used to update previous research results. By utilizing this method and its expected technology, vulnerability studies and mapping can be done more quickly, practically, easily, and efficiently, especially if additional or updated variables exist. Finally, the results of this study can be considered when planning and managing coastal areas to address the challenges posed by global climate change.

Data and methods

This study's stages are divided into 6 (six) steps, as illustrated in Fig. 2. These stages encompass accessing geospatial data from multiple sources and categorizing the data; normalizing and stacking the layers of data; creating the training data set (which involves training, consuming, and

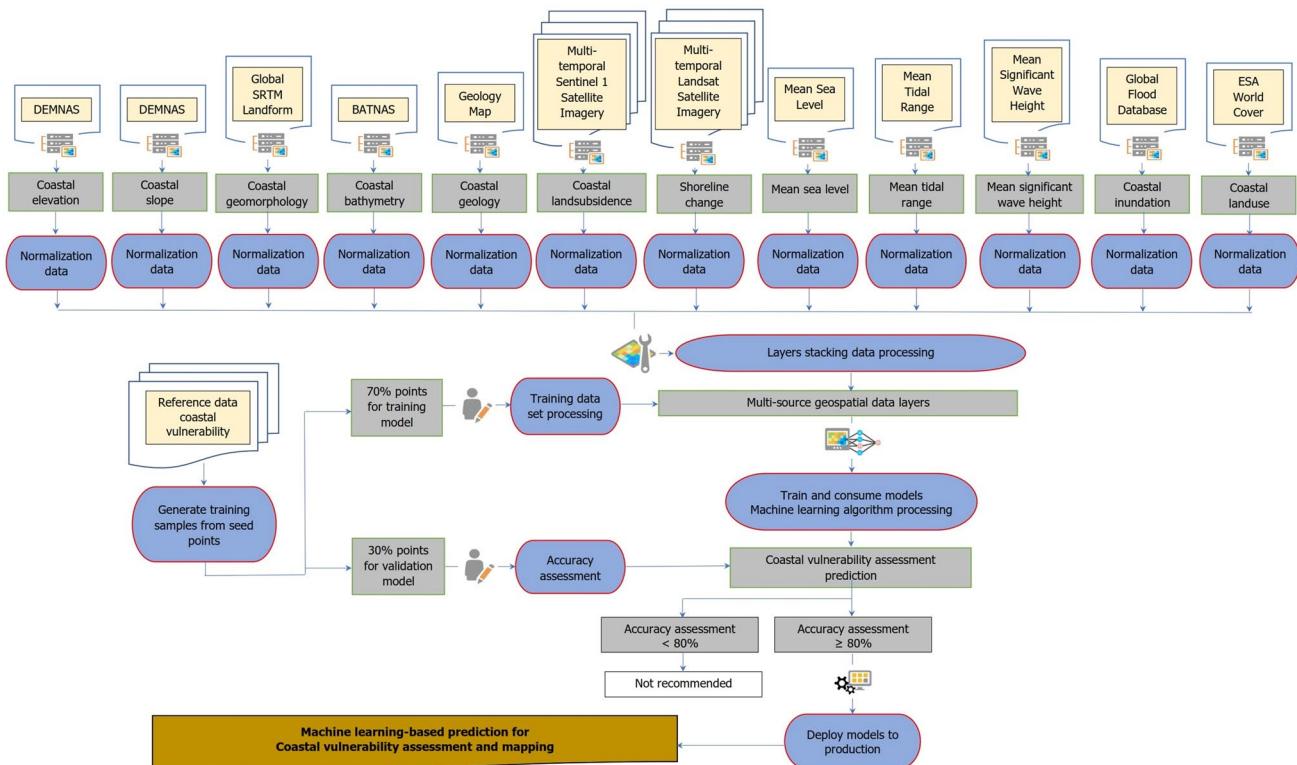


Fig. 2 The flow chart in this study is divided into 6 (six) research stages

optimizing the machine learning model); conducting accuracy assessment; and deploying the models for the production of CVA and mapping. Detailed explanations of the data and methods used in this study are provided in the following sub-chapters.

Datasets

14 (fourteen) sources of multi-source geospatial data were accessed and applied in this study, as shown in Table 1. Subsequently, this multi-source geospatial data was categorized, merged, and utilized as input variables for machine learning tree-based algorithms used in CVA and mapping within the area of interest. The variables used in merging and grouping the multi-geospatial data in this study are presented in Table 2. For reference, this study used CVI data from 2009 provided by the Ministry of Marine and Fisheries, Republic

of Indonesia, at a 1:250,000 scale, as well as CVI data from previous studies (e.g., Anwar et al. 2020; Handiani et al. 2022). The dataset includes training data for model training and validation data for assessing model accuracy. This study focuses on the geo-bio-physical aspects of CVA variables, which were adopted from the CVI Gornitz model (Gornitz 1990). These variables encompass mean tidal range, coastal elevation, slope, geomorphology, land use, bathymetry, geology, land subsidence, shoreline change, mean sea level, mean significant wave height, and flood inundation.

Coastal elevation

Elevation becomes a significant factor in coastal vulnerability assessment (Parthasarathy and Natesan 2015; Hossain et al. 2022; Handiani et al. 2022). Coastal elevation is used to identify and estimate the areas along the coast that are

Table 1 Multi-source geospatial data used in this study

Data	Source	Spatial resolution	Temporal Resolution	Reference
DEMNAS	Geospatial Information Agency (BIG), Republic of Indonesia	8 m	2018	https://tanahair.indonesia.go.id/
BATNAS	Geospatial Information Agency (BIG), Republic of Indonesia	180 m	2018	https://tanahair.indonesia.go.id/
Global SRTM Landforms	Conservation Science Partners	90 m	2006–2011	Theobald et al. (2015)
ESA World Cover	ESA/VITO/Brockmann Consult/CS/GAMMA Remote Sensing/IIASA/WUR	10 m	2020–2021	Zanaga et al. (2021)
Annual Landsat 5 TM satellite imagery	USGS	30 m	1990, 1995	https://www.usgs.gov/landsat-missions/landsat-collection-2-level-2-science-products
Annual Landsat 7 ETM + satellite imagery	USGS	30 m	2000, 2005, 2010	
Annual Landsat 8 OLI/TIRS satellite imagery	USGS	30 m	2015, 2020	
Sentinel 1 satellite imagery	European Union/ESA/Copernicus	20 m	2015–2020	https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/
Global Flood Database	Cloud to Street (C2S) / Dartmouth Flood Observatory (DFO)	250 m	2000–2018	Tellman et al. (2021)
Mean Sea Level	Copernicus Marine Environment Monitoring Service	27.5 km	1993–2021	https://www.aviso.altimetry.fr/en-data/products/ocean-indicators-products/mean-sea-level.html
Mean Significant Wave Height	Copernicus Marine Environment Monitoring Service	9 km	2020–2021	https://data.marine.copernicus.eu/product/GLOBAL_ANALYSIS_FORECAST_WAV_001_027/services
Mean Tidal Range	Ministry of Marine and Fisheries, Republic of Indonesia	1:250,000	2009	https://www.kkp.go.id/ http://pusriskel.litbang.kkp.go.id/index.php/en/peta-kerentanan-pesisir-nasional
Geology map	ESDM	1:250,000	1993	https://onemap.esdm.go.id/map/geologi.html
Coastal vulnerability index	Ministry of Marine and Fisheries, Republic of Indonesia	1:250,000	2009	https://www.kkp.go.id/ http://pusriskel.litbang.kkp.go.id/index.php/en/publikasi

Table 2 The grouping variables of multi-source geospatial data are used in this study

No	Data	Coastal elevation	Coastal slope	Coastal bathymetry	Coastal geo-morphology	Coastal landuse	Coastal inundation	Coastal shoreline subsidence	Coastal geology	Mean Sea Level	Mean Tidal Range	Mean Significant Wave Height	Reference
1	DEMNAS	✓	✓	-	-	-	-	-	-	-	-	-	-
2	BATNAS	-	-	✓	-	-	-	-	-	-	-	-	-
3	Global SRTM Landforms	-	-	-	✓	-	-	-	-	-	-	-	-
4	ESA World Cover	-	-	-	-	✓	-	-	-	-	-	-	-
5	Annual Landsat 5 TM satellite imagery	-	-	-	-	-	-	-	-	-	-	-	-
6	Annual Landsat 7 ETM + satellite imagery	-	-	-	-	-	-	✓	-	-	-	-	-
7	Annual Landsat 8 OLI/TIRS satellite imagery	-	-	-	-	-	-	✓	-	-	-	-	-
8	Sentinel 1 satellite imagery	-	-	-	-	-	-	-	✓	-	-	-	-
9	Global Flood Database	-	-	-	-	-	-	✓	-	-	-	-	-
10	Mean Sea Level	-	-	-	-	-	-	-	-	-	-	-	-
11	Mean Significant Wave Height	-	-	-	-	-	-	-	-	-	-	✓	-
12	Mean Tidal Range	-	-	-	-	-	-	-	-	-	-	-	-
13	Geology map	-	-	-	-	-	-	-	-	-	-	-	-
14	Coastal vulnerability index	-	-	-	-	-	-	-	-	-	-	✓	-

exposed to future sea level rise. Coastal elevation represents the mean elevation of a specific area above the mean sea level. High-elevation coastal areas exhibit greater resilience to inundation caused by storm surges, rising sea levels, and tsunami run-ups. Meanwhile, coastal areas are considered vulnerable when elevations are low (Beluru Jana and Hegde 2016). The elevation data for this study were obtained from the Digital Elevation Model National (DEMNAS), which has an 8-m spatial resolution (The Geospatial Information Agency (BIG) 2018, Source: <https://tanahair.indonesia.go.id/>).

Coastal slope

The coastal slope is a crucial factor in assessing coastal vulnerability, and numerous previous studies have considered this variable (Roukounis and Tsirhrintzis 2022). Coastal slope represents the difference in elevation between two points over a horizontal distance along the coast. These data play a significant role in determining vulnerability to floods in coastal areas, as areas with lower slopes are more susceptible to erosion than those with steeper slopes. The coastal slope is also used to assess relative vulnerability to inundation and shoreline changes (Thieler and Hammar-Klose 1999; Mani Murali et al. 2013). In this study, coastal slope data are obtained from DEMNAS using the algorithm described in Eq. (1).

$$CS = ee.Terrain.slope(input) \quad (1)$$

CS is the coastal slope in degrees, and $input$ is the data for an elevation image from DEMNAS in meters. $ee.Terrain.slope$ is the module function to create slope data in GEE. The local gradient is calculated using the 4-connected neighbors of each pixel that occur around the edges of an image (Source: <https://developers.google.com/earth-engine/apidocs/ee-terrain-slope>).

Coastal geomorphology

Geomorphology is a scientific discipline that studies the processes, evolution, and landforms of the Earth's surface. Coastal morphology is shaped by factors such as tectonics, geological structure, lithology, and erosion processes (Mani Murali et al. 2013). Geomorphological variables can describe the relative erodibility of different landforms (Ermini et al. 2005; Metelka et al. 2018). Rocky cliffs and benches offer maximum resistance to waves, thereby reducing coastal vulnerability, while soft and muddy features such as low dunes, sabkha, mudflats, and mangroves provide minimal resistance (Rao et al. 2009). In this study, coastal geomorphology data were obtained from Global SRTM Landforms. Conservation Sciences Partners produced this

global SRTM Landforms data in 2006 – 2011, with a spatial resolution of 90 m (Theobald et al. 2015).

Coastal bathymetry

Bathymetry measures ocean floor depths relative to sea level (Dierssen and Theberge 2016). Coastal bathymetric variable for assessing their impact on coastal vulnerability (Sankari et al. 2015). Bathymetric conditions are crucial in influencing erosion and accretion processes, making them essential for determining coastal vulnerability, encompassing erosion impact and accretion processes (Hossain et al. 2022; Kabiri 2017; Hossain et al. 2022). In addition, bathymetric depth levels directly impact the extent of inundation in coastal areas (Parthasarathy and Natesan 2015). Greater seawater penetration and shallower bathymetry increase the risk in coastal areas (Hedger et al. 2001). In this study, coastal bathymetry data were obtained from the National Bathymetry (BATNAS) dataset, which offers a spatial resolution of 180 m (The Geospatial Information Agency (BIG) 2018, Source: <https://tanahair.indonesia.go.id/>).

Coastal geology

In this study, coastal geology refers to lithological information identifying rock types in coastal systems. These data variables are essential for understanding the resistance of the coastal substrate to erosion, a characteristic determined by the relative hardness of the minerals in the rock type (Gornitz 1990; Remondo et al. 2003). Coastal areas with resistant lithology are less vulnerable to erosion caused by rising sea levels and storm surges. Meanwhile, coastal areas with unconsolidated sediments are highly susceptible to erosion (Radiarta et al. 2022). According to Handiani et al. (2022), bedrock and coastal constituent materials, such as sand, mud, or rocky plains, also significantly impact coastal vulnerability. In this study, coastal geology data were obtained from a geological map in 1993, at a scale of 1: 250,000, provided by the Ministry of Energy and Mineral Resources, Republic of Indonesia (Source: <https://onemap.esdm.go.id/map/geologi.html>).

Coastal land subsidence

Coastal land subsidence refers to the gradual sinking of land surfaces in the coastal area, resulting from natural or human-induced processes. A high rate of land subsidence serves as an indicator of vulnerability in coastal areas (Addo 2013). Balica et al. (2012) and Addo (2013) have utilized this variable to assess vulnerability in coastal cities, recognizing its significance. The combined effects of sea level rise and land subsidence pose substantial risks to society, ecological systems, and the economy (Triana and Wahyudi 2020). Land subsidence can increase the vulnerability to disasters,

including tidal floods, which frequently impact the northern coastal areas of Java, Indonesia (e.g., Marfai and King 2007; Bott et al. 2021). Furthermore, recent studies have used Sentinel-1 SAR data for subsidence analysis (Dong et al. 2021; Nasiri et al. 2021; Sidiq et al. 2021). This study's land subsidence data were derived from SAR Sentinel-1 multi-temporal data from 2015 to 2020. These data, accessible free of charge, have a spatial resolution of 20 m and are provided by the European Space Agency/Copernicus. The information on land subsidence was generated using the Small Baseline Subset Interferometric Synthetic Aperture Radar (SBAS-InSAR) method (e.g., Manunta et al. 2019; Cigna and Tapete 2021; Anjasmara et al. 2020). The SBAS-InSAR represents an advancement of the Differential Interferometric Synthetic Aperture Radar (DInSAR) technique (Berardino et al. 2002; Manunta et al. 2019). This method leverages phase differences between two SAR data acquisitions with varying timestamps over the same geographical area, accurately providing spatial information about deformation (Li et al. 2022). Moreover, the SBAS method offers advantages over the DInSAR method by minimizing atmospheric propagation disturbances, topographical interference, and temporal decorrelation (Berardino et al. 2002).

Shoreline change

Shoreline changes result from natural and human activities, driven by erosion and sedimentation processes influenced by wave characteristics, nearshore circulation, sediment characteristics, and beach morphology (Mani Murali et al. 2013). Various factors contribute to shoreline change, including river sediment supply, alongshore currents, sea-level rise, and storm surge (Hossain et al. 2022). High shoreline change rates, primarily due to erosion, render coastal areas vulnerable to environmental shifts (Parthasarathy and Natesan 2015). As a result, shoreline changes serve as essential geological variables in numerous CVA studies (Mahamoud et al. 2022). This study analyzed shoreline changes using Multi-temporal Landsat satellite data from 1990 to 2020, obtained from the US Geological Survey (USGS) with a spatial resolution of 30 m. Annual Landsat 5 data with TM sensors were used to determine shoreline data in 1990 and 1995, while annual Landsat 7 data with ETM+ sensors were used for shoreline data in 2000, 2005, and 2010. Annual Landsat 8 data with OLI/TIRS sensors were employed for shoreline data in 2015 and 2020. The data collected served as input for rate-of-change statistics calculations using the Digital Shoreline Analysis System (DSAS), an automatic tool that computes statistical measures of shoreline change from multiple and historical shoreline positions (Himmelstoss et al. 2018; Gómez-Pazo et al. 2022). Furthermore, DSAS has been widely used in shoreline change analysis in various studies (e.g., Rao et al. 2009; Parthasarathy and

Natesan 2015; Beluru Jana and Hegde 2016; De Serio et al. 2018; Hossain et al. 2022; Kumar et al. 2022). DSAS can also be applied to predict changes in lake shorelines (Wu et al. 2022). Within DSAS, several statistical parameters, including end point rate, linear regression rate, shoreline change, and net shoreline movement, are computed to measure the rates of shoreline displacement (erosion or accretion) (Dereli and Tercan 2020). In this study, we used the results of rate-of-change statistics calculations, specifically the Linear Regression Rate (LRR) in meters per year, to describe coastal shoreline conditions between 1990 and 2020.

Mean sea level

Mean Sea Level (MSL) is a tidal datum defined as the arithmetic mean hourly surface elevation measured over a 19-year cycle. MSL provides a reference point for the average sea level height relative to local coastal benchmarks during specific timeframes, such as a year or a month (Pramanik et al. 2016). On the other hand, sea level rise represents the long-term increase in the average sea level of the oceans, measured from the center of the Earth's mass, the sea floor, or relative to sea level. It significantly impacts MSL (Stammer et al. 2012). Changes in ocean volume drive global sea level rise due to factors such as melting ice and thermal expansion (Mani Murali et al. 2013), and it is a critical indicator of our climate's health (Haddad et al. 2013). Sea level fluctuations result from various factors, including global conditions, local conditions, and physical features, leading to seasonal variations (De Serio et al. 2018). Therefore, sea level rise is among the variables used to assess coastal vulnerability, along with average tidal range and significant wave height (Lin et al. 2020). In this study, MSL data for 1993 to 2021 were obtained from the Copernicus Marine Environment Monitoring Service, offering a spatial resolution of 27.5 km (Source: <https://www.aviso.altimetry.fr/en/data/products/ocean-indicators-products/mean-sea-level.html>).

Mean tidal range

Mean Tidal Range (MTR) represents the difference between the lowest and highest tide levels and is typically measured in meters (Mani Murali et al. 2013). Coastal areas characterized by high tidal ranges are generally considered highly vulnerable, while those with low tidal ranges are viewed as less vulnerable (Rao et al. 2009; Hossain et al. 2022; Handiani et al. 2022). The length of the tidal range plays a crucial role in determining the spatial extent of the coast affected by waves, as intertidal relief zones with minimal tidal range are particularly susceptible to permanent inundation from sea-level rise (Doukakis 2005). For this study, MTR data were obtained from the 2009 dataset provided by KKP, using a 1:250,000 scale.

Mean significant wave height

The Mean Significant Wave Height (MSWH) represents the deviation between the trough and crest of waves, serving as a crucial indicator of wave energy that plays a significant role in CVA by influencing coastal sediments (Pendleton et al. 2005). Waves are a hydrodynamic factor generated by energy transfer through water in coastal areas (Rao et al. 2009; De Serio et al. 2018). The wave height directly affects the volume of beach material that can be transported offshore and permanently removed from the coastal sediment system (Doukakis 2005). Coastal areas with high wave energy are typically classified as highly vulnerable, while those with low wave energy are considered less vulnerable (Hossain et al. 2022). In this study, data for the MSWH for 2020 to 2021, with a spatial resolution of 9 km, were sourced from the Copernicus Marine Environment Monitoring Services.

Coastal inundation

Coastal inundation is characterized by flooding coastal areas due to increased river flows, rising sea levels, storms, and the impact of tsunami waves (Gayathri et al. 2017). Several studies conducted by Sahoo and Bhaskaran (2018) and Sudha et al. (2015) have employed this variable to assess vulnerability in coastal areas. For this study, the Global Flood Database obtained coastal inundation data from 2000 to 2018, with a spatial resolution of 250 m. This dataset was published by Cloud to Street (C2S) / Dartmouth Flood Observatory (DFO), offering a comprehensive record of 913 flood events occurring between 2000 and 2018. These flood events were meticulously collected from the Dartmouth Flood Observatory and utilized to acquire MODIS imagery from Terra and Aqua sensors (Tellman et al. 2021). Additionally, several other studies have relied on this dataset, including Demirkesen et al. (2008), Marfai and King (2008), Balica et al. (2012), Wu (2021), and Barros et al. (2022).

Coastal land use

Due to the climate change phenomenon, there is a need for land use data to monitor changes in environmental and coastal vulnerability (Jamali 2020). Land use significantly impacts coastal vulnerability, either increasing or decreasing it (Hossain et al. 2022). Moreover, one of the socio-economic aspects must be considered for CVA (Mani Murali et al. 2013). Land use refers to using land for various purposes, such as recreation, tourism, agriculture, residence, and the ground's surface cover, including vegetation, water bodies, urban infrastructure, bare soil, and others (De Serio et al. 2018). This study obtained coastal land use from ESA World Cover for 2020 to 2021, published by ESA/VITO/Brockmann Consult/CS/GAMMA Remote Sensing/IIASA/

WUR and has a spatial resolution of 30 m (Zanaga et al. 2021).

Normalization and layer stacking data processing

Data normalization is a data processing technique aimed at rescaling or transforming input data features to bring them into the same scale range. This study utilized normalization to enhance performance, optimization, and stability during the training of models. Furthermore, this method is invaluable in mitigating issues that can reduce accuracy during training models with outliers and dominant feature data (Singh and Singh 2020). Additionally, prior studies conducted by Sola and Sevilla (1997) have demonstrated that normalization significantly reduces errors and expedites computation time during the model training process. This study applied the Min–Max normalization method to normalize the input training data. This method employs a linear transformation to convert the input data into new data with the same interval value. The normalized data falls within a value range between 0 and 1. The Min–Max normalization method can be presented by Eq. (2) (Pandey and Jain 2017).

$$X_{\text{new}} = (x - \min(x)) / (\max(x) - \min(x)) \quad (2)$$

Where: X_{new} is the value of data sets after normalization. x is the value of data sets before normalization. \min is the minimum value of the data sets. \max is the maximum value of the data sets.

Layer stacking is combining separate image variables to create new multi-image variables. Several image variables must have the same level, number of rows, and columns to perform layer stacking (Bui and Mucsi 2022). This study employs this process to merge and integrate multiple variables from multi-source geospatial data. Furthermore, the resulting composite of geospatial, created through layers stacking, serves as input for processing the training dataset, training the machine learning model, and generating predictions for CVA.

Training data set processing

The CVI data provided by the KKP and CVI data from previous studies were utilized in this study as reference data. The reference dataset was selected using the systematic sampling method outlined by Bornstein et al. (2013), considering a point-to-point distance of less than 2.5 km and coastal vulnerability classes from reference data sources. These data were processed to generate the processing data set, including training and validation datasets for model validation.

The dataset was divided into two parts, following the standard 70 and 30 splitting ratio, with 70% of the total area allocated for training the machine learning models (training

dataset) and the remaining 30% used for model validation and accuracy assessment. While the 70/30 splitting ratio is widely used in many cases, it is not a rigid rule. Depending on the complexity and size of the data, some tasks may require different proportions (Muraina 2022). These studies also observed that most academics suggest a 70 and 30 proportion (70% for training and 30% for testing) for small dataset sizes, typically ranging from one hundred to one million data points. This splitting ratio of 70 and 30 has also been recommended by other researchers, such as Nguyen et al. (2021). It is advantageous in achieving optimal model performance due to sufficient data for training and testing purposes. However, selecting an appropriate splitting ratio ensures model stability and better testing performance (Birba 2020). In this study, machine learning algorithms for CVA with an 80% or higher accuracy result are recommended for use as CVA production models. Conversely, deploying machine learning algorithms for production is not advisable if the accuracy assessment results fall below 80%.

Train, consume, and optimize the model of machine learning algorithms processing

This study employed machine learning tree-based algorithms within the GEE environment for CVA prediction and mapping. The GEE platform facilitated the comparison and evaluation models during training to achieve the highest performance and accuracy in predicting and mapping coastal vulnerability. The tree-based algorithms utilized in this study include RF, GTB, and CART. Tree-based algorithms are well-suited for handling smaller datasets and are known for their precision and accuracy, outperforming other machine learning algorithms. RF, in particular, demonstrates high precision and accuracy regardless of the number of attributes and data instances (Akinsola 2017).

Additionally, measuring feature importance is crucial for determining the significance of variables in each machine-learning input. A standard method for assessing feature importance in tree-base models is the mean decrease in impurity, aggregating the improvements associated with each note split to calculate the importance (e.g., Noi Phan et al. 2020). Optimizing the machine learning learning process is essential to achieve the most optimal accuracy assessment for each algorithm's setup parameters.

Classification and regression trees

One of the machine learning techniques for constructing predictive models from datasets is Classification and Regression Trees (CART). The CART is generated by recursively classifying datasets and applying a straightforward predictive model without developing prediction equations. The result is displayed as decision three, visually representing

the clustering process. The classification tree is designed for the dependent variable, considering a limited number of discrete values that can be measured based on misclassification. The calculation typically involves computing the squared variance between the predicted and observed values to measure the prediction error due to misclassification (Loh et al. 2019). The data is divided into subsets with homogeneous values for the dependent variable along the predictor axis. Furthermore, the decision tree illustrates the process employed to generate a prediction based on new observations (Krzywinski and Altman 2017).

Gradient tree boost

Gradient Tree Boost (GTB) is a machine learning technique that generates predictive models in the form of grouped models based on decision trees to address classification and regression problems (Fafalios et al. 2020). GTB has gained widespread popularity due to its high accuracy, fast training, quick prediction times, and efficient memory usage. The GTB algorithm can advance multitasking in machine learning, particularly for multi-class classification tasks (Ke et al. 2017). Consequently, numerous studies have been conducted using the GTB algorithm, including those by Zheng et al. (2008), Yingli et al. (2020), Adler and Painsky (2022).

Random forest

The Random Forest (RF) algorithm is a supervised machine-learning technique that can be employed to construct and combine multiple decision trees. This algorithm functions by selecting the majority class for classification tasks and can also be utilized for regression by computing the average output of all the trees. In this algorithm, individual decision trees may produce errors. However, when aggregated as a group, they are more accurate, allowing for the generation of a more reliable overall decision. As a result, several studies have been conducted using the RF algorithm, including those by Tian et al. (2016), Praticò et al. (2021), Zhao et al. (2022).

Performance model assessment

The highest level of performance and accuracy is achieved by analyzing tree-based machine learning models to predict coastal vulnerability in the study area. The accuracy tuning process involves varying machine learning parameters and systematically comparing and evaluating the results at each step (e.g., Husnayaen et al. 2018; Abijith and Saravanan 2021). The Overall Accuracy (OA) calculations are employed to assess the accuracy of various tree-based machine-learning algorithms. This accuracy assessment relies on reference training datasets, constituting 30% of the

total datasets. All algorithms were executed with varying numbers of trees to attain the optimal accuracy assessment in this study. The number of trees was incremented by 10, ranging from 10 to 1000, resulting in 100 models for each algorithm in this accuracy assessment. As a result, an overall accuracy of at least 80% for the prediction model using tree-based machine-learning algorithms is recommended. It can be applied to the modeling and mapping of CVA in the study area.

The OA parameter evaluates the model's performance based on the evaluation matrix. The OA of the classification model is calculated using Eq. (3) (Grandini et al. 2020).

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Where OA is the overall accuracy, TP is the True positive, TN is the True Negative, FP is the False Positive, and FN is the False Negative.

Furthermore, the Kappa index (K-Index) assesses model performance. K-Index is a statistical method used to measure the degree of agreement between the classification model and the actual label (e.g., Talukdar et al. 2020; Martínez Prentice et al. 2021; Thanh Noi and Kappas 2018). According to Cohen (1960), the K-Index is expressed using Eq. (4):

$$Ki = \frac{Po - Pe}{1 - Pe} \quad (4)$$

Where Ki is the K-index, Po is the proportion of true classification. Pe is the expected proportion of cases classified truly.

Model deployment

The final stage of this study involves implementing the model to predict CVA in the study area. The trained model is fed with 12 coastal vulnerability variables extracted from multi-sources during this stage, as mentioned in Table 1. The entire model implementation process was conducted with the GEE platform.

Results

Layers stacking and merging multi-source geospatial data

This study utilized 14 multi-source geospatial data from various sources. These datasets were integrated, and a layers-stacking data processing technique was applied to create input variables for machine learning, enabling the study area's prediction and mapping of coastal vulnerability. These parameters include coastal elevation, MTR, slope,

geomorphology, bathymetry, geology, land subsidence, land use, MSL, shoreline change, MSWH, and flood inundation. The CVA variables were grouped and integrated, as shown in Figs. 3, 4 and 5, with the figure illustrating the western, middle, and eastern parts of the study area, respectively.

Figure 3 describes CVA variables in the western part of the study area, encompassing the Serang, Jakarta, Karawang, and Indramayu Regions. The coastal elevation variable is presented in Fig. 3a, revealing that most coastal areas have low elevations and gentle coastal slopes, as presented in Fig. 3b. Figure 3c indicates that the dominant area has a shallow depth (< -2.87 m). The coastal geomorphology variable (Fig. 3d) shows a lower slope and elevation prevalence in the study area. The land use variable is depicted in Fig. 3e, including cropland, open water, built-up, and mangroves. Figure 3f displays the MSL variable, showcasing varying values. Specifically, the area around Indramayu features values ranging from 0.82 – 0.96 m/year, while Serang and Jakarta area values range from 0.53—0.82 m/year. Figure 3g reveals the shoreline change variable, with an average of -13.03 – 20.36 m/year. Figure 3h demonstrates the frequency of flood inundation variable, with an average of 4—> 11 events/year.

The MTR variable is presented in Fig. 3i, with an average value ranging from 0.8 – 1.6 m. Figure 3j displays the MSWH variable ranging from 0.29 to > 0.73 m. The coastal land subsidence variable (Fig. 3k) indicates rates between 0.01 – 0.08 m/year, with higher rates observed in the Serang and Jakarta areas. The geological variable (Fig. 3l) shows that the western part is predominantly composed of sedimentary rocks, which are more susceptible to erosion. Figure 4 represents CVA variable data in the middle, including the Cirebon, Tegal, Pekalongan, Kendal, Semarang, and Demak to Jepara regions. The coastal elevation variable (Fig. 4a) reveals low elevation with a gentle 0 – 2.3 degrees slope in Fig. 4b. The coastal bathymetry (Fig. 4c) indicates a dominant shallow depth (<-2.87 m). The coastal geomorphology variable (Fig. 4d) displays lower and upper slopes with low elevations. Figure 4e illustrates diverse land use, encompassing cropland, open water, built-up, and mangroves.

Figure 4f, i, h describe MSL, shoreline change, flood inundation, and MRT variables, respectively. Regarding MSL, the Cirebon area exhibits values from 0.68 – 0.96 m/year, while Tegal, Pekalongan, and Semarang show higher values exceeding 0.96 m/year. The shoreline changes in this area range from -55.96 – 20.36 m/year, with Brebes, Semarang, Demak, and Jepara experiencing the highest rates. The flood inundation variable indicates frequent occurrences, averaging 4 to 11 events/ year, especially in Cirebon, Semarang, Demak, and Jepara. MTR exhibits an average value of approximately 1.2 – 1.6 m.

Figure 4j, k, and l represent the MSWH, land subsidence, and geological variable, respectively. The MSWH

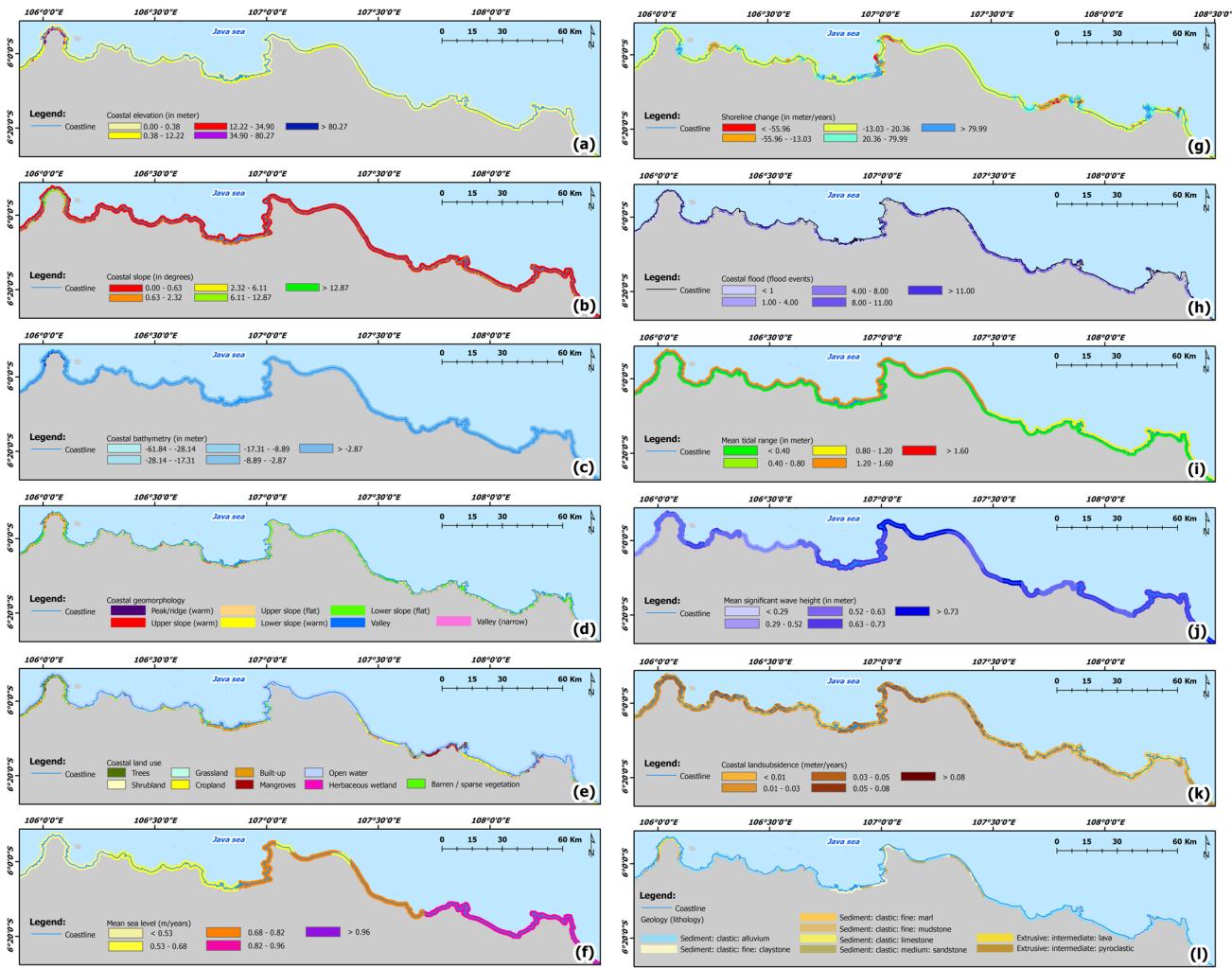


Fig. 3 The results of grouping and integrating variables data for the study area in the western part. **a** coastal elevation, **b** slope, **c** bathymetry, **d** geomorphology, **e** land use, **f** mean sea level, **g** shoreline

change, **h** flood inundation, **i** mean tidal range, **j** mean significant wave height, **k** land subsidence, **l** geology

in this region range from 0.52—> 0.73 m. The coastal land subsidence rate varies from 0.01 – 0.08 m/year, with higher rates observed in Cirebon, Pemalang, Batang, and Kendal. The geological variable indicates a dominance of sedimentary rocks in the area. Figure 5 presents the integrated variable data in the eastern part, including Pati, Rembang, Tuban, Lamongan, and Gresik regions. Figure 5a and Fig. 5b depict low elevations (0 – 34.9 m) with a gentle slope (0 – 6.11 degrees). The bathymetry (Fig. 5c) indicates shallow water with a depth of -2.78 m. Coastal geomorphology variables (Fig. 5d) reveal that the mountains and hills around Mount Muria influence the upper slope. Land use diversity is showcased in Fig. 5e, featuring cropland, open water, built-up, and mangroves. In this area, the MSL data ranges from 0.68 to 0.96 m/year (Fig. 5f). The rate shoreline changes (Fig. 5g)

exhibit rates between -13.03 and 20.36 m/year, while the flood events variable (Fig. 5h) frequently affect the eastern area, with an average of 1–11 events/year. Pati and Gresik regions experience the highest frequency of 8–11 events/year.

The MTR variable in the eastern part (Fig. 5i) averages around 0.8 to > 1.6 m, with Lamongan and Gresik regions recording the highest values. Additional variables, including MSWH (Fig. 5j), coastal land subsidence (Fig. 5k), and geological composition (Fig. 5l), are also presented. The MSWH ranges from 0.29 to 0.73 m, while coastal subsidence rates vary from 0.01 to > 0.08 m/year, with Rembang and Tuban exhibiting higher rates (> 0.08 m/year). Geological analysis indicates a prevalence of sedimentary rocks in this area, which are more susceptible to erosion than other rock types.

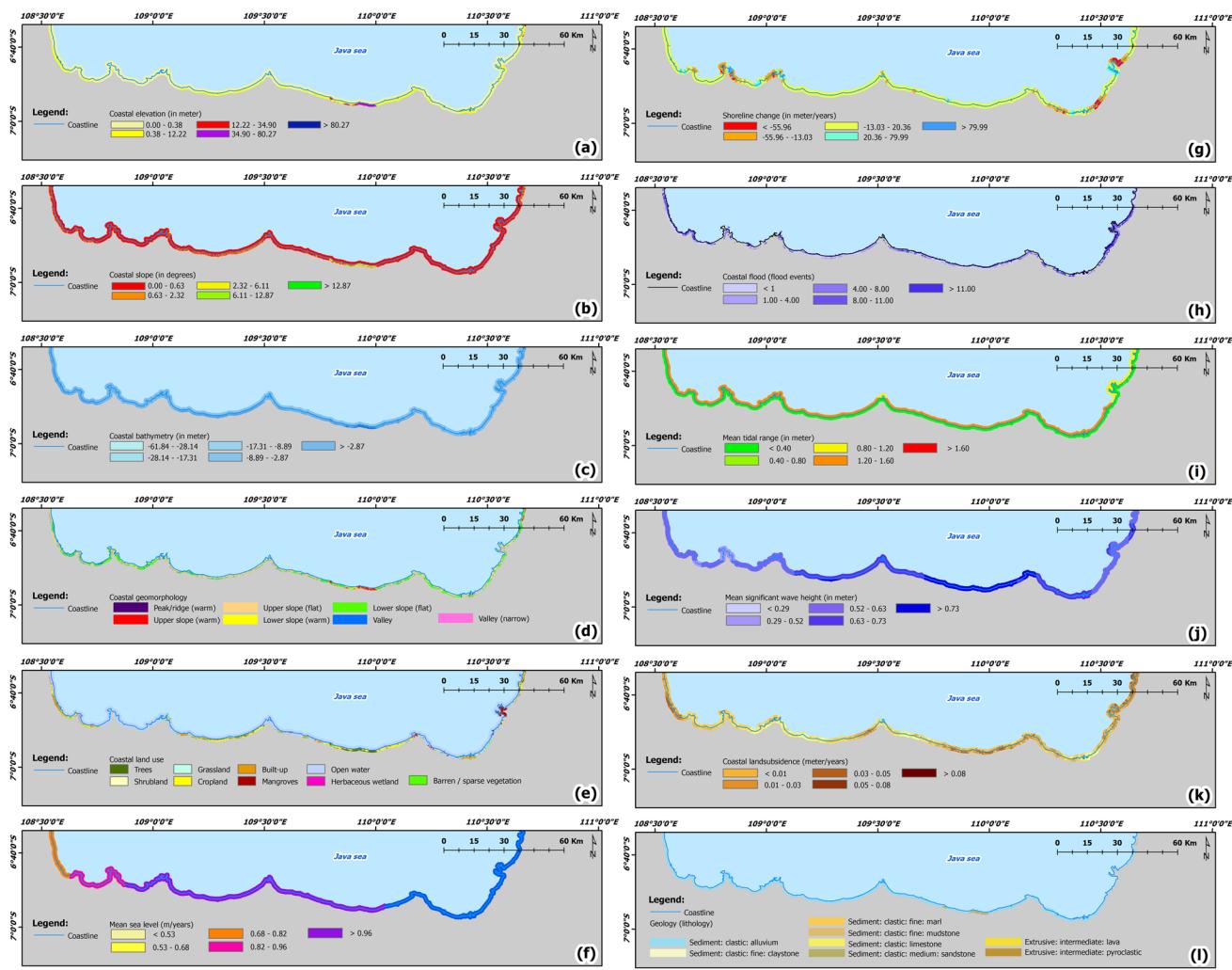


Fig. 4 The results of grouping and integrating variables data for the study area in the middle part. **a** coastal elevation, **b** slope, **c** bathymetry, **d** geomorphology, **e** land use, **f** mean sea level, **g** shoreline

change, **h** flood inundation, **i** mean tidal range, **j** mean significant wave height, **k** land subsidence, **l** geology

Relative importance variables machine learning tree-based algorithms modeling for CVA predictions and mapping

Figure 6 compares the relative importance of variables in CVA for the CART, GTB, and RF algorithms. The order of importance for each variable varies among the machine learning algorithms used.

For the RF algorithm, the most crucial variables are bathymetry, slope, and shoreline change. Conversely, the GTB algorithm prioritizes mean sea level, slope, and mean significant height as the most important variables. Meanwhile, the CART algorithm assigns the highest importance to elevation and slope. Two variables, mean tidal range and land subsidence, are of lesser importance in the CVA according to the CART algorithm. However, some variables exhibit significantly higher importance than others, such as

elevation in the CART algorithm and the MSL in the GTB algorithm.

CART Algorithm is a classification method based on a recursive binary splitting dataset, meaning the datasets are partitioned into two different parts and refer to the threshold value (Lewis 2000). The partitioning processes are repeated until the model meets the satisfied classification according to class data in the dataset. These splitting processes consider a single variable; therefore, the CART algorithm will produce better performance if trained with one of the variables or simple variables combination (Orieschnig et al. 2021). In Fig. 6, the Elevation parameter accounted as the highest (the most important feature). It indicates that the elevation data has a significant influence and has become the main reference in recursive binary splitting in CVA machine learning with the CART algorithm. As a result, the trained CVA model with the CART algorithm is highly sensitive to

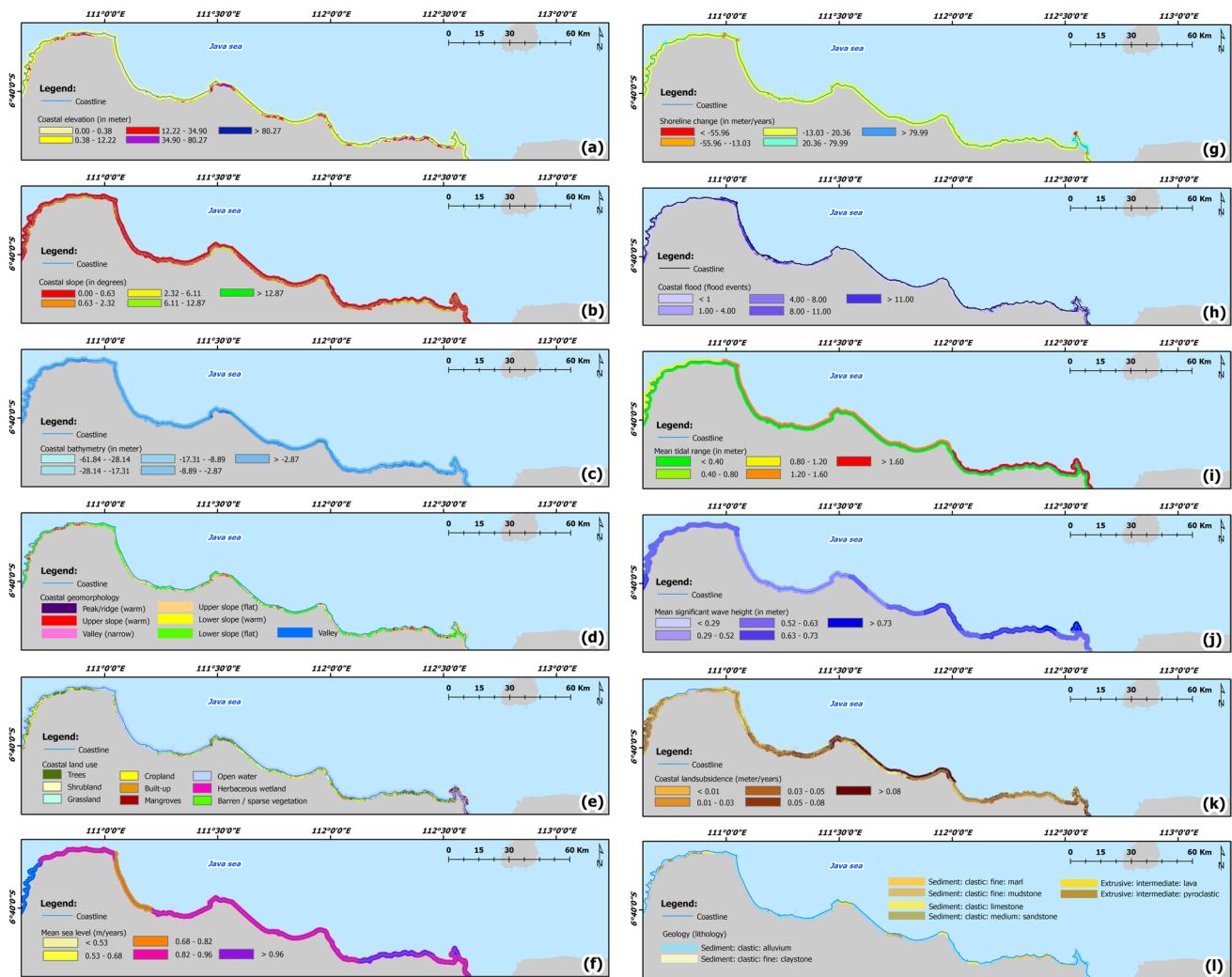


Fig. 5 The results of grouping and integrating variables data for the study area in the eastern part. **a** coastal elevation, **b** slope, **c** bathymetry, **d** geomorphology, **e** land use, **f** mean sea level, **g** shoreline

change, **h** flood inundation, **i** mean tidal range, **j** mean significant wave height, **k** land subsidence, **l** geology

changes in elevation data and leads to different tree structures, which may cause miss classification.

The results of this model indicate that additional variables beyond those in the Gornitz model hold relatively high relative importance, namely elevation, land subsidence, and bathymetry. Meanwhile, geology and land use variables appear to have lower relative importance.

Spatial data distribution of CVA

The distribution of CVA generated by the machine learning tree-based algorithms for the study area in the western, middle, and eastern parts is presented in Figs. 7, 8 and 9, respectively. Each figure illustrates the CVA assessment result

using the RF, GTB, and CART algorithms. The Coastal vulnerability is classified into 5 (five) classes: very low, low, moderate, high, and very high. Figure 7 and Table 3 depict the coastal vulnerability in the western part. According to the CART algorithm (first sub-figure), the vulnerability in this area is predominantly classified as high class (345 km or 58.6%). Regions such as Subang, Karawang, and a small part of Jakarta fall into the very high class.

Similarly, based on the GTB algorithm, the vulnerability in this area is also mostly classified as high class (315 km or 53.4%), with only a small area in Jakarta falling into the very high class. In contrast, according to the RF algorithm, vulnerability is predominantly in the very high class (478% or 81.1%), with small areas in Jakarta

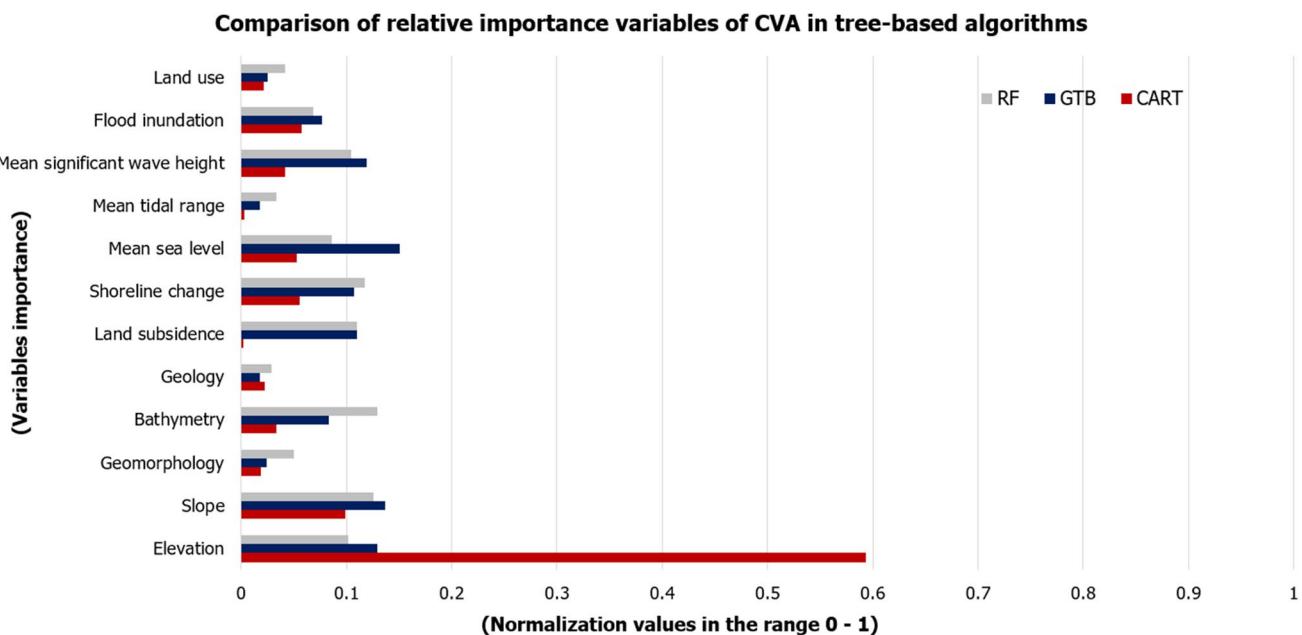


Fig. 6 The results of the comparison of relative importance variables of CVA for the algorithms of CART, GTB, and RF

and Indramayu falling into the very high CVA class. Figure 8 and Table 4 show the results of CVA's spatial distribution for the middle study area. According to the CART algorithm, vulnerability in this area is mainly in the very high class (288 km or 64.7%), encompassing regions such as Brebes, Tegal, Pekalongan, Kendal, and Demak. Based on the GTB algorithm, the vulnerability is predominantly in the high class (212 km or 47.7%), with locations like Brebes, Tegal, Pekalongan, Demak, and Jepara classified as very high. Meanwhile, the RF algorithm shows that vulnerability in the middle part is dominated by the high (226 km or 50.8%) and very high (203.8 km or 45.8%) classes, with regions including Brebes, Tegal, Pekalongan, Semarang, Kudus, and Jepara classified in the very high class.

Figure 9 and Table 5 display the results for the eastern part. The majority of the vulnerability in this area is classified as very high (173 km or 37.2%) and high class (148 km or 31.7%) based on the CART algorithm, with locations like Tuban, Lamongan, and Gresik falling into the very high class. According to the GTB algorithm, the vulnerability in the eastern part is dominated by the moderate class (103 km or 22.2%) and high class (261 km or 56%). In particular, Tuban and Gresik are classified in the very high class. Meanwhile, based on the RF algorithm, most of the eastern part is classified as high class (293 km or 62.8%), with Gresik being classified in the very high CVA class. In summary, vulnerability along the north coast of Java is generally classified as high and very high.

Discussion

The following discussion will demonstrate how the CVA model's comparisons and conclusions relate to other study's limitations and potential for future applications.

Accuracy, consistency, comparisons with other studies and reference data

Figures 7, 8 and 9 present the final result of CVA based on the tree-based algorithms for the study area's western, middle, and eastern parts, respectively. The comparisons are made between the CVA prediction results from the tree-based algorithms with the CVI reference data in 2009 from the Ministry of Marine and Fisheries, Republic of Indonesia, as well as CVI reference data from previous studies (e.g., Anwar et al. 2020; Handiani et al. 2022). Our reference dataset consists of 538 sample points, with 406 points (70% of the reference data) used for model training and 177 points (30%) for model validation.

Previous studies of coastal vulnerability in the western part provided results that align with this research. Anwar et al. (2020) stated that the vulnerability on the West Java coast is high, especially in the northern region. Handiani et al. (2022), using the CVIw method, showed that coastal vulnerability in Subang-West Java is 43%, classified as very high. For the middle part, the results of this study

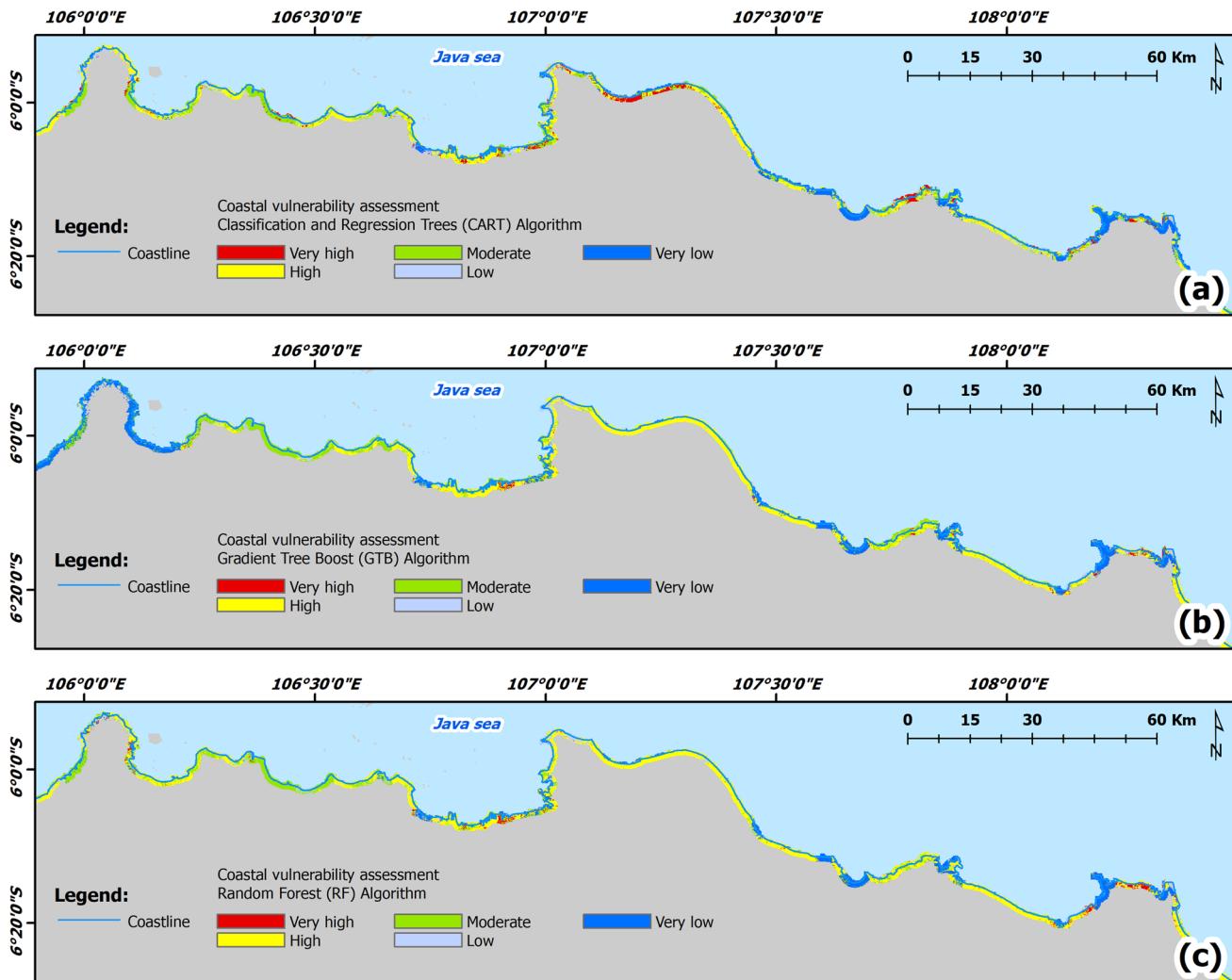


Fig. 7 The results of spatial distribution and comparison of CVA from the performance of machine learning tree-based algorithms modeling for the study area in the western part. **a** CVA from the CART algorithm, **b** CVA from the GTB algorithm, **c** CVA from the RF algorithm

are similar to several previous studies. Prasetia (2021) reported that 51.43% of the coastal area on the north coast of Central Java is vulnerable. Bengen and Tahir (2012) stated that coastal vulnerability in Pekalongan is classified as very high (4.76%) and high (42.86%). Bengen and Tahir (2012) also examined coastal vulnerability in Surabaya-East Java, and the results showed that 35.56% of Surabaya's beaches were classified as vulnerable and 15.56% as very vulnerable, consistent with the research results classifying Surabaya as high-risk. Handiani et al. (2022) also assessed coastal vulnerability in northern Java using the CVIw method. The results of this research indicate that vulnerability on the north coast of Java is predominantly high (39%) and very high (51%), which is in line with the results of this research.

Figure 10 displays the optimization results for the tree-based algorithms used in CVA. The most optimal accuracy assessments and consistencies are achieved with the following parameters: CART, GTB, and RF algorithms are 71.18% with 120 numbers of trees, 77.40% with 650 numbers of trees, and 80.22% with 20 numbers of trees. In summary, Fig. 10 demonstrates that the RF algorithm exhibits better accuracy assessment trends than CART and GTB algorithms.

While the CART and GTB algorithms exhibit low accuracy, their accuracy assessments continue to increase with the number of trees until they stagnate at 120 and 650, respectively. The RF algorithm's higher accuracy is attributed to its ability to reduce overfitting without significantly increasing bias-induced errors. The RF algorithm

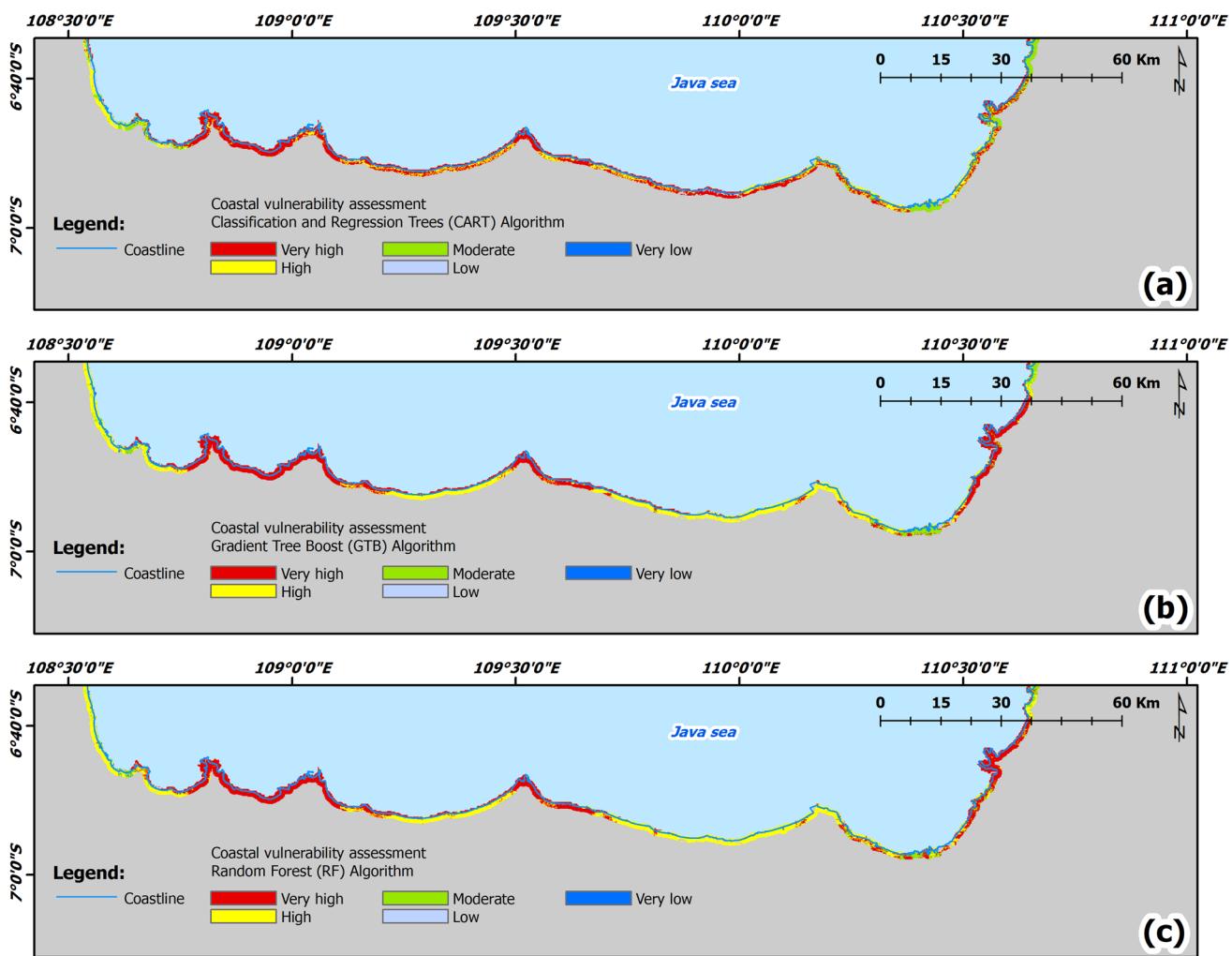


Fig. 8 The results of spatial distribution and comparison of CVA from the performance of machine learning tree-based algorithms modeling for the study area in the middle part. **a** CVA from the CART algorithm, **b** CVA from the GTB algorithm, **c** CVA from the RF algorithm

combines numerous decision trees, leading to optimal results (Montesinos López et al. 2022; Ao et al. 2018). Moreover, prior studies have also reported that the RF algorithm outperforms other algorithm's accuracy (e.g., Noi Phan et al. 2020; Cha et al. 2021; Praticò et al. 2021).

This study conducted Google Street View searches to gain insights into the field conditions corresponding to each vulnerability class. Google Street View provides 360-degree panoramic views of specified streets within its coverage area, mirroring the coverage area of Google Maps. Figures 11, 12 and 13 present the results, including photos of CVA conditions from Google Street View.

Figure 11 provides examples of photos and field conditions in the western part of the study area obtained from Google Street View, related to CVA classes resulting from the RF algorithm. The photo captions in Fig. 11 represent

field conditions for different CVA classes, including (a) CVA in high class, (b) CVA in moderate class, (c) CVA in high class, (d) CVA in low class, (e) CVA in low class, (f) CVA in very low class.

Similarly, Fig. 12 presents photos and field conditions in the central part of the study area obtained from Google Street View, corresponding to CVA classes resulting from the RF. The photo captions in Fig. 12 depict field conditions for different CVA classes, including (a) CVA in high class, (b) CVA in very high class, (c) CVA in high class, (d) CVA in very high class, (e) CVA in high class, (f) CVA in very high class.

Figure 13 displays examples of photos and field conditions in the eastern part of the study area from Google Street, linked to CVA classes resulting from the RF algorithm. The Photo captions in Fig. 13 illustrate field

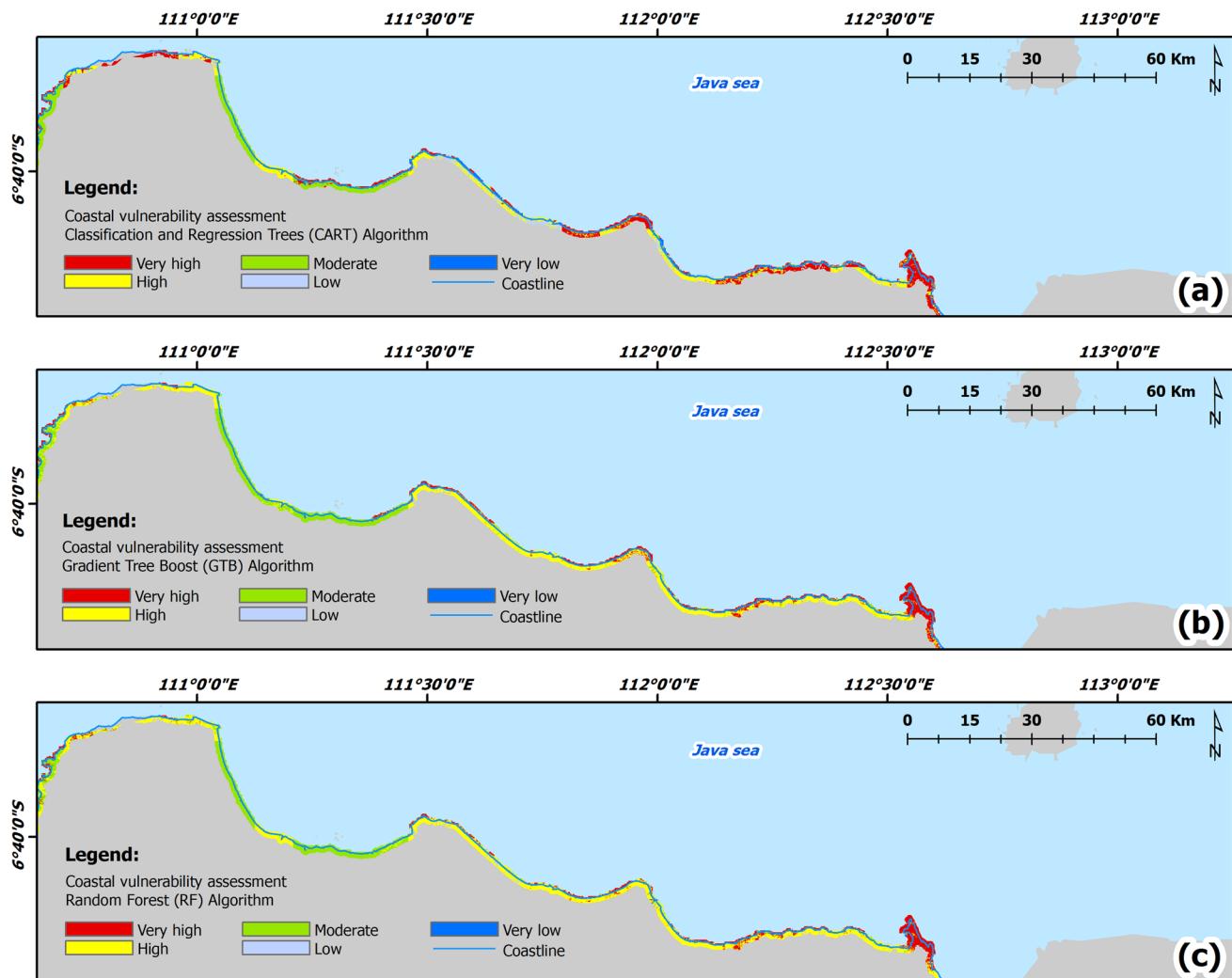


Fig. 9 The results of spatial distribution and comparison of CVA from the performance of machine learning tree-based algorithms modeling for the study area in the eastern part. **a** CVA from the CART algorithm, **b** CVA from the GTB algorithm, **c** CVA from the RF algorithm

Table 3 The comparison results of shoreline length in CVA class from the performance of machine learning tree-based algorithms approach for the study area in the western part

Class CVA	RF		CART		GTB	
	km	%	km	%	km	%
Very low	44.9	7.6	106.1	18.0	153.2	26.0
Low	0.0	0.0	22.6	3.8	22.6	3.8
Moderate	66.0	11.2	102.1	17.3	98.0	16.6
High	0.0	0.0	345.7	58.6	315.3	53.4
Very high	478.3	81.1	12.7	2.1	0.0	0.0
Total	590.0	100.0	590.0	100.0	590.0	100.0

conditions for different CVA classes, including (a) CVA in high class, (b) CVA in moderate class, (c) CVA in high class, (d) CVA in high class, (e) CVA in moderate class, (f) CVA in low class.

Based on these photos (Figs. 11–13), areas classified as very high and high vulnerability appear damaged, dirty, and poorly maintained, while those with lower vulnerability exhibit better conditions.

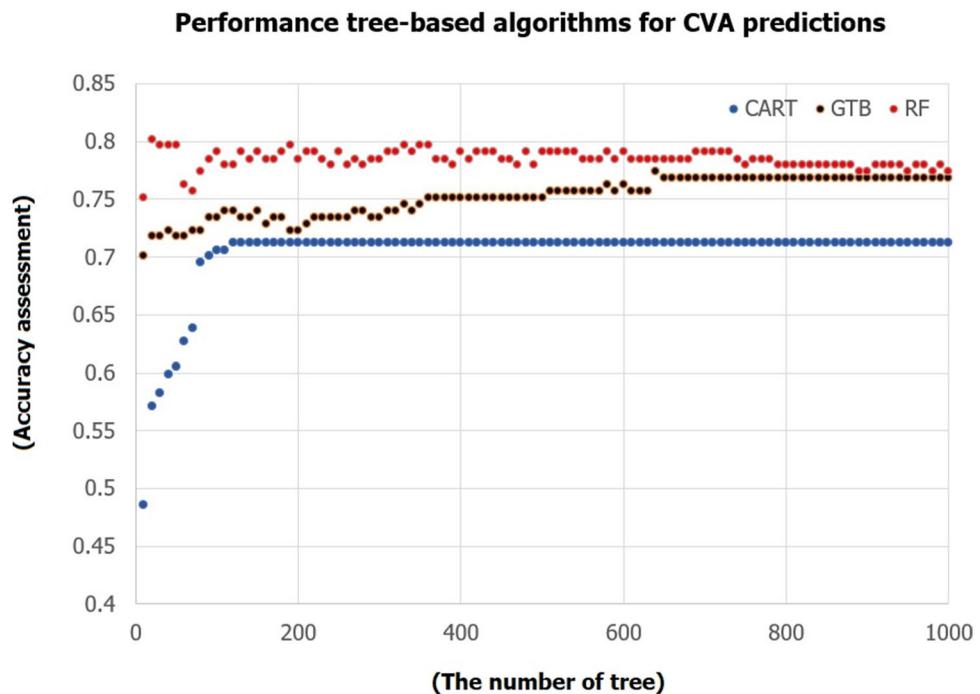
Table 4 The comparison results of shoreline length in CVA class from the performance of machine learning tree-based algorithms approach for the study area in the middle part

Class CVA	RF		CART		GTB	
	km	%	km	%	km	%
Very low	0.0	0.0	0.0	0.0	0.0	0.0
Low	0.0	0.0	0.0	0.0	0.0	0.0
Moderate	15.3	3.4	89.5	20.1	49.5	11.1
High	226.0	50.8	67.5	15.2	212.1	47.7
Very high	203.8	45.8	288.1	64.7	183.5	41.2
Total	445.0	100.0	445.0	100.0	445.0	100.0

Table 5 The comparison results of shoreline length in CVA class from the performance of machine learning tree-based algorithms approach for the study area in the eastern part

Class CVA	RF		CART		GTB	
	km	%	km	%	km	%
Very low	27.5	5.9	27.5	5.9	27.5	5.9
Low	6.1	1.3	16.7	3.6	6.1	1.3
Moderate	94.6	20.3	100.7	21.6	103.9	22.2
High	293.5	62.8	148.2	31.7	261.6	56.0
Very high	45.3	9.7	173.9	37.2	61.1	13.1
Total	467.0	100.0	467.0	100.0	467.0	100.0

Fig. 10 The results of performance machine learning-based predictions for CVA in the study area



Overall accuracy and kappa index correlation

Overall Accuracy (OA) and Kappa Index (K-Index) are popular parameters to measure machine learning performance (e.g., Pham et al. 2018; Qi et al. 2018; Taalab et al. 2018). Table 6 shows the OA and K-Index of the CVA

machine learning algorithm. Based on Table 6, the OA and K-Index for each CVA machine learning type demonstrate relatively good performance. The OA is notably high and is complemented by a moderate K-Index level for CART and a substantial K-Index level for RF and GTB. These K-Index classification levels align with the research

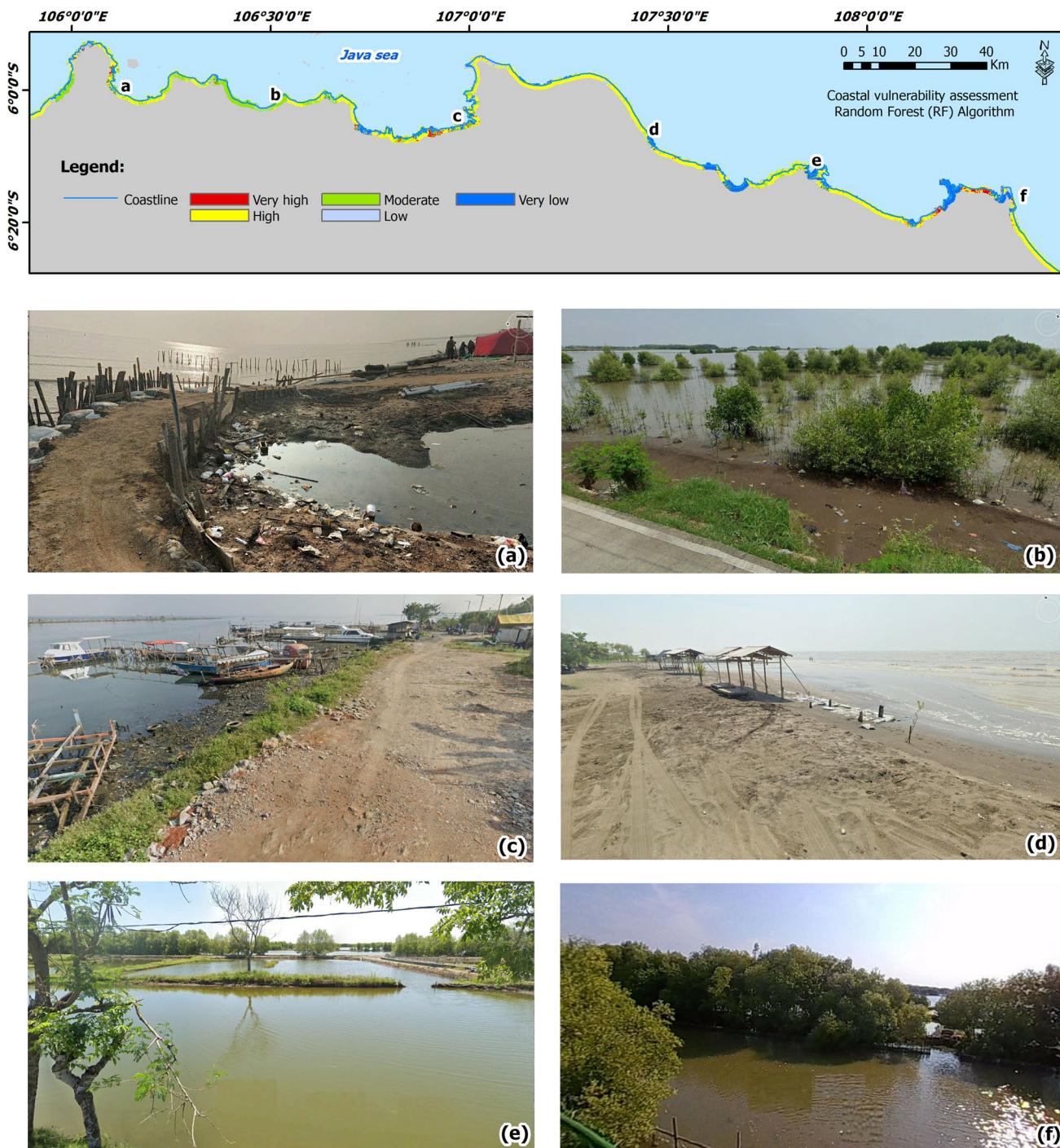


Fig. 11 Example of photos and field conditions in the western part study area from Google Street View (google maps) related to the CVA class resulting from the Random Forest algorithm. Photo Cap-

tions (a) CVA in high class, b CVA in moderate class, c CVA in high class, d CVA in low class, e CVA in low class, f CVA in very low class

conducted by Landis and Koch (1977) and Sim and Wright (2005), which categorized the K-Index into six levels: poor, slight, fair, moderate, substantial, and almost perfect.

The high OA indicates that the model can recognize patterns within the dataset used for training and testing.

It is the model that performs just as well when predicting new data. However, it is important to note that this parameter can lead to biased interpretations if the class distribution in the dataset is not balanced. Meanwhile, the K-Index represents a satisfactory result, with values exceeding 0.50

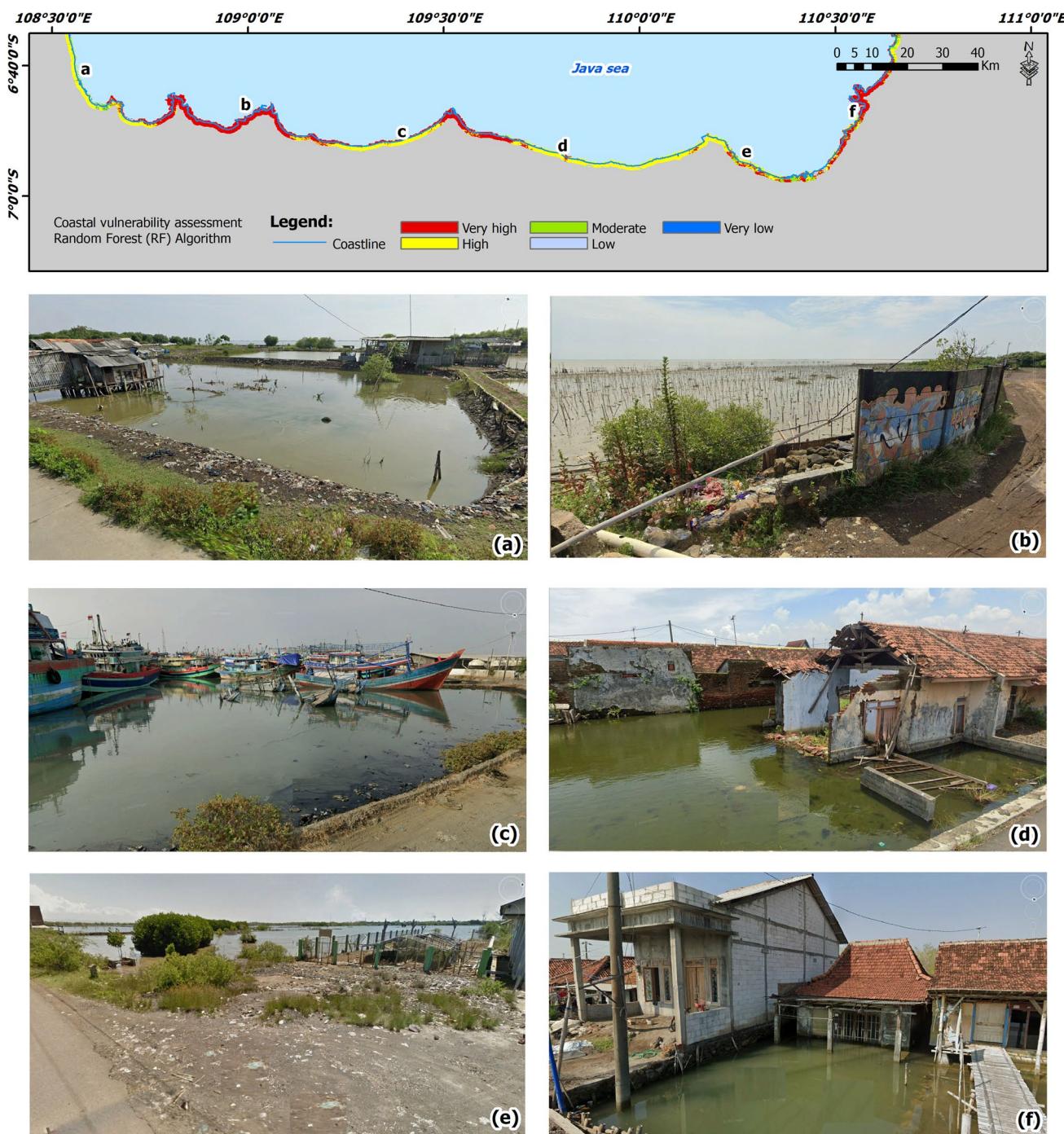


Fig. 12 Examples of photos and field conditions in the middle part of the study area from Google Street View (google maps) related to the CVA class resulting from the Random Forest algorithm. Photo Cap-

tions (a) CVA in high class, b CVA in very high class, c CVA in high class d CVA in very high class, e CVA in high class, f CVA in very high class

for all machine learning algorithms. It implies that the model exhibits positive performance and achieves a higher level of agreement than expected by chance. The correlation between OA and K-Index can be observed in Table 6. Generally, the OA values are greater than the K-Index,

which suggests that the models may have been trained with an imbalanced class dataset.

Consequently, these models are more adept at recognizing certain classes of data, particularly the dominant ones. To clarify the class distribution, we have calculated the number

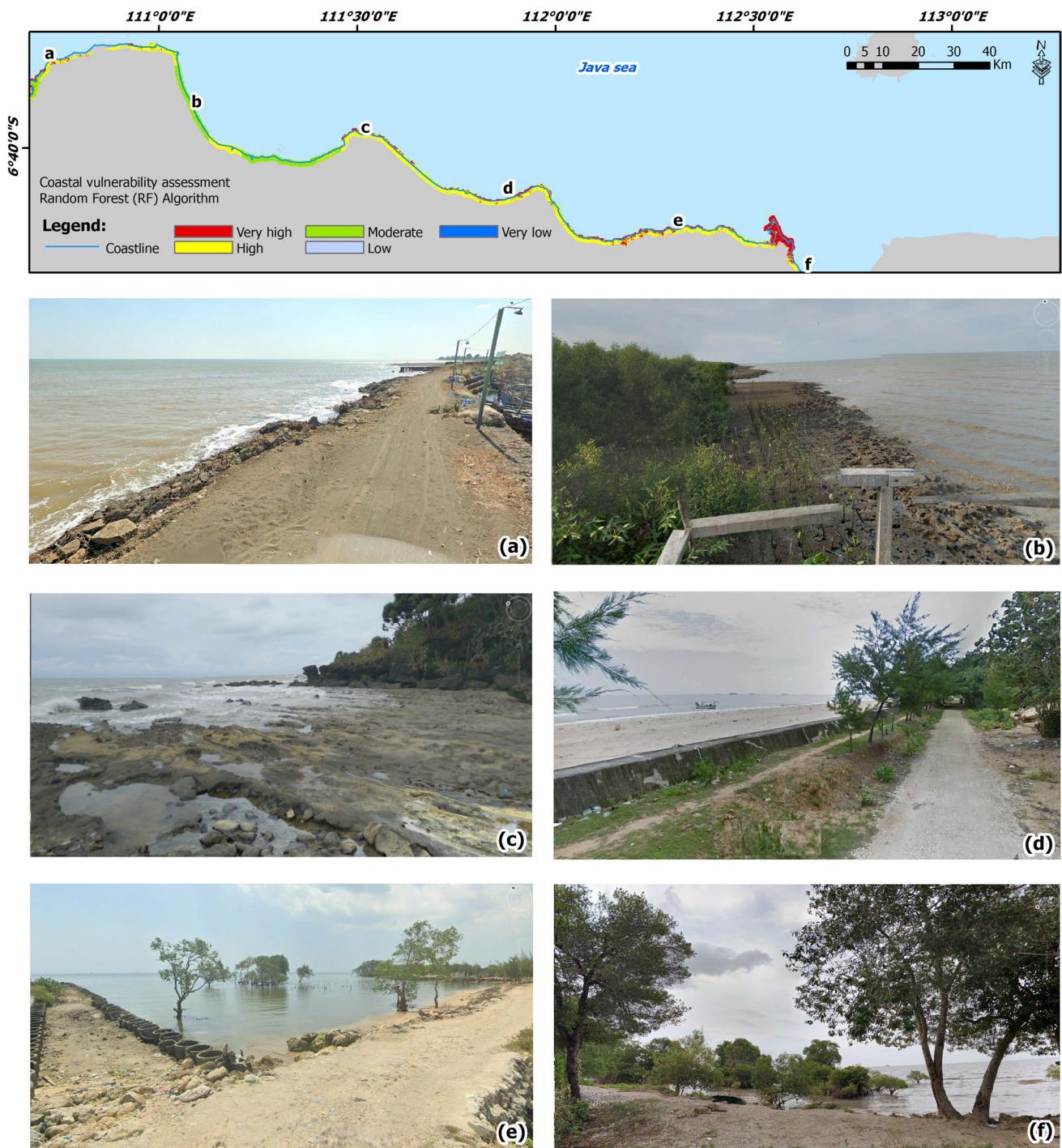


Fig. 13 Example of photos and field conditions in the eastern part study area from Google Street View (google maps) related to the CVA class resulting from the Random Forest algorithm. Photo Cap-

tions (a) CVA in high class, b CVA in moderate class, c CVA in high class, d CVA in high class, e CVA in moderate class, f CVA in low class

of members for each class in the CVA dataset. The very high, high, and moderate classes account for 133, 174, and

99 points, respectively. In contrast, the low and very low classes have 50 and 52 points as members.

Table 6 Overall Accuracy and Kappa Index of CVA machine learning

CVA Machine Learning Types	OA	K-Index
RF	0.80	0.72
GTB	0.77	0.67
CART	0.71	0.58

The class distribution fluctuates, indicating that the data classes are not evenly distributed. The very high, high, and moderate classes are more dominant than the other classes. Increasing the number of datasets and balancing the class distribution may be considered to address this issue.

Research implication and recommendations

According to Hossain et al. (2022) and Aliño et al. (2013), CVA can map coastal vulnerability to waves. CVA emphasizes the role of ecosystems in coastal protection from flood, regulating ecosystem services, assisting decision-makers, determining priority areas of vulnerability, addressing environmental problems, protecting coastal ecosystems from damage, and evaluating vulnerability related to coastal disasters.

Numerous studies have been developed using various methods to create CVA models. These methods can be categorized into four types: indexes, indicators, GIS, and dynamic computer models (Noor and Abdul Maulud 2022). The CVA model obtained based on the index and indicator approach is relatively easy to implement. However, incorporating new parameters and updating existing ones can be time-consuming and labor-intensive (e.g., Blasiak et al. 2017; Debortoli et al. 2019; Giannakidou et al. 2020).

Pradeep et al. (2022) utilized a K-means clustering machine learning algorithm to predict shoreline change along the coast between Pozhiyoor and Anchuthengu, India. This study employed extensive machine learning modeling to assess and map coastal vulnerability by incorporating 12 parameters. This study used a tree-based algorithm (RF, CART, and GTB). By leveraging this technology, the assessment and mapping of vulnerabilities can become more efficient. The implication is that assessing, mapping, and updating coastal vulnerabilities can be accomplished more quickly, affordably, easily, and with greater resource efficiency.

The findings of this study are expected to provide valuable insights for decision-makers involved in coastal disaster management in the study area, given the complexity of coastal issues. We also hope this method can be adopted for studying, mapping, and regularly updating the vulnerability levels of all beaches in Indonesia in collaboration with the Ministry of Maritime Affairs and Fisheries of the Republic

of Indonesia. The results of this study can offer valuable insights to the central government and local authorities, highlighting specific locations requiring special attention. By doing so, governments can prioritize and strategize coastal disaster adaptation and mitigation efforts, thereby safeguarding the coastal environment, including its geo-bio-physical and socio-economic aspects, at an early stage. This approach minimizes potential risks and promotes sustainable and integrated coastal area management.

Upcoming studies and limitations

This study has limitations related to variations in spatial and temporal resolutions among the input variables derived from multi-source geospatial data, which may introduce bias to the results. Future research should focus on developing methods to extract geospatial data with consistent spatial and temporal resolutions to mitigate such bias and align with accurate evaluation results. The CVA reference data used in this study were obtained from the Ministry of Marine and Fisheries (KKP), the Republic of Indonesia, and previous studies due to the unavailability of direct field measurements (e.g., Anwar et al. 2020; Handiani et al. 2022).. While these references are currently the best available, future research could benefit from direct field observations and measurements to improve accuracy. Additionally, the CVA mapping achieved in this study is suitable for scales ranging from 1:100,000 to 1:250,000.

Further studies may consider mapping CVA at a medium to large scale of 1:5,000 to 1:50,000. This study employed the GEE platforms CART, GTB, and RF algorithms. Future research could explore alternative algorithms like Naive Bayes and Neural Networks for improved CVA accuracy. Enhancing the dataset by increasing data volume and balancing class distribution is another area for improvement.

Conclusions

The CVA and mapping can be predicted effectively using 12 parameters and the tree-based algorithms. The OA on the RF, GTB, and CART algorithms was 80.22%, 77.40%, and 71.18%, respectively. The RF algorithm demonstrates better accuracy trends compared to CART and GTB. The K-index of the RF algorithm is 0.72, indicating that this model performs adequately in classifying the data. The importance level of each parameter varies for each algorithm type. In general, vulnerability along the north coast of Java is classified as high and very high, consistent with field conditions. Future research may explore using other machine learning algorithms, such as Naive Bayes and Neural Networks, based on reference data obtained from field surveys.

Author contributions F.Y, M.W the main contributor, conceived and designed the experiment; F.Y, M.W, D.H.F.P, E.A.W, Y.P, A.R performed and analyzed machine learning experiments; F.Y, M.W, A.Y pre-processed the database and analyzed data on beach elevation, slope, geomorphology and geology; F.Y, M.W, A.Y pre-processed databases and analyzed data on flood inundation, land use, and land subsidence; N.R, A.N, S.N pre-processed the database and analyzed the shoreline change data; M.Y.I, E.R, A.S.P pre-processed the database and analyzed data on mean sea level, average tidal range, mean significant wave height; H.A, I.F performed database pre-processing and analyzed bathymetry data; F.Y, M.W, A.Y performed and analyzed CVA and mapping; I.F, S.N performed and analyzed research recommendations; All authors wrote the paper; All authors have read and agree to the published version of the manuscript.

Funding This study was supported by the internal research program of the Research Center for Hydrodynamics Technology, National Research and Innovation Agency (BRIN), Indonesia in 2022.

Data availability The authors can confirm that all relevant data this study uses is included in the manuscript's article and information files. The data were available on request from the authors, and availability of data can be accessed free of charge from the data source. The datasets generated during and analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

Conflict of Interest/Competing Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Abidin HZ, Andreas H, Gumilar I et al (2013) Land subsidence in coastal city of Semarang (Indonesia): characteristics, impacts and causes. *Geomat Nat Haz Risk* 4:226–240. <https://doi.org/10.1080/19475705.2012.692336>
- Abijith D, Saravanan S (2021) Assessment of land use and land cover change detection and prediction using remote sensing and CA Markov in the northern coastal districts of Tamil Nadu, India. *Environ Sci Pollut Res*. <https://doi.org/10.1007/s11356-021-15782-6>
- Achiari H, Wulandari N, Yustiani YM, Harlan D (2015) Study erosion and coastal destruction at Pondok-Bali, North Coast-West Java of Indonesia. *Int J Manag Appl Sci* 1:317–320
- Addo KA (2013) Assessing coastal vulnerability index to climate change: the case of Accra – Ghana. *J Coastal Res* 165:1892–1897. <https://doi.org/10.2112/SI65-320.1>
- Adler AI, Painsky A (2022) Feature importance in gradient boosting trees with cross-validation feature selection. *Entropy* 24. <https://doi.org/10.3390/e24050687>
- Akinsola JET (2017) Supervised machine learning algorithms: classification and comparison. *Int J Comput Trends Technol* 4(3):128–138
- Aliño PM, Follosco NMG, Mamauag SS, Martinez RJS, Panga FM (2013) Vulnerability assessment tools for coastal ecosystems - a guidebook. *Mar Environ Resour Found* 7:1–164
- Anjasmaria IM, Yulyta SA, Taufik M (2020) Application of time series InSAR (SBAS) method using sentinel-1A data for land subsidence detection in Surabaya city. *Int J Adv Sci Eng Inf Technol* 10:191–197. <https://doi.org/10.18517/ijaseit.10.1.6749>
- Anwar SK, Purba NP, Yuniarini, Subiyanto (2020) Coastal Vulnerability Based on Oceanographic and Ecosystem Parameters on the North and South Coast of West Java. In: Proceeding - AGERS 2020: IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology: Understanding the Interaction of Land, Ocean and Atmosphere: Disaster Mitigation and Regional Resilience. IEEE, pp 184–190. <https://doi.org/10.1109/AGERS51788.2020.9452761>
- Arjasakusuma S, Kusuma SS, Saringatin S et al (2021) Shoreline dynamics in East Java Province, Indonesia, from 2000 to 2019 using multi-sensor remote sensing data. *Land* 10:1–17. <https://doi.org/10.3390/land10020100>
- Astsatryan H, Grigoryan H, Abrahamyan R et al (2022) Shoreline delineation service: using an earth observation data cube and sentinel 2 images for coastal monitoring. *Earth Sci Inf* 15:1587–1596. <https://doi.org/10.1007/s12145-022-00806-7>
- Atikawati D-, Gunawan T, Sunarto S (2019) Penerapan Etika Lingkungan Dalam Pengelolaan Wilayah Kepesisiran Tuban. *Jurnal Geografi Geografi dan Pengajarannya* 17:1. <https://doi.org/10.26740/jgg.v17n1.p1-10>
- Bagheri M, Ibrahim ZZ, Akhir MF et al (2021) Impacts of Future Sea-level rise under global warming assessed from tide gauge records: a case study of the East Coast economic region of Peninsular Malaysia. *Land* 10:1382. <https://doi.org/10.3390/LAND10121382>
- Balica SF, Wright NG, van der Meulen FA (2012) Flood vulnerability index for coastal cities and its use in assessing climate change impacts. *Nat Hazards* 64:73–105. <https://doi.org/10.1007/s11069-012-0234-1>
- Barman NK, Chatterjee S, Paul AK (2016) Estimate the coastal vulnerability in the Balasore Coast of India: a statistical approach. *Model Earth Syst Environ* 2:20. <https://doi.org/10.1007/s40808-015-0074-6>
- Barros JL, Tavares AO, Santos PP, Freire P (2022) Enhancing a coastal territorial vulnerability index: anticipating the impacts of coastal flooding with a local scale approach. *Coast Manag* 50:442–468. <https://doi.org/10.1080/08920753.2022.2107858>
- Beluru Jana A, Hegde AV (2016) GIS Based Approach for Vulnerability Assessment of the Karnataka Coast, India. *Adv Civil Eng* 2016. <https://doi.org/10.1155/2016/564253>
- Bengen DG, Tahir A (2012) Policy review: opportunities for enhancing coastal community resilience and climate change adaptation in Indonesia. United States Agency for International Development. Bogor. https://www.crc.uri.edu/download/IMC_CCPolPaperFinal_11_12.pdf
- Berardino P, Fornaro G, Lanari R, Sansosti E (2002) A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Trans Geosci Remote Sens* 40:2375–2383. <https://doi.org/10.1109/TGRS.2002.803792>
- Blasiak R, Spijkers J, Tokunaga K et al (2017) Climate change and marine fisheries: least developed countries top global index of vulnerability. *PLoS ONE* 12:e0179632. <https://doi.org/10.1371/JOURNAL.PONE.0179632>
- Bornstein MH, Jager J, Putnick DL (2013) Sampling in developmental science: situations, shortcomings, solutions, and standards. *Dev Rev* 33(4):357–370. <https://doi.org/10.1016/j.dr.2013.08.003>
- Bott LM, Schöne T, Illigner J et al (2021) Land subsidence in Jakarta and Semarang Bay – The relationship between physical processes, risk perception, and household adaptation. *Ocean Coast Manag* 211. <https://doi.org/10.1016/j.ocecoaman.2021.105775>
- Bui DH, Mucsi L (2022) Comparison of layer-stacking and Dempster-Shafer theory-based methods using Sentinel-1 and Sentinel-2 data fusion in urban land cover mapping. *Geospatial Inf Sci*. <https://doi.org/10.1080/10095020.2022.2035656>
- Bukvic A, Rohat G, Apotsos A, de Sherbinin A (2020) A systematic review of coastal vulnerability mapping. *Sustainability* (Switzerland) 12:1–26. <https://doi.org/10.3390/su12072822>

- Cha GW, Moon HJ, Kim YC (2021) Comparison of random forest and gradient boosting machine models for predicting demolition waste based on small datasets and categorical variables. *Int J Environ Res Public Health* 18. <https://doi.org/10.3390/ijerph18168530>
- Chakraborty S (2021) Remote Sensing and GIS in Environmental Management. *Environmental Management: Issues and Concerns in Developing Countries* 185–220. https://doi.org/10.1007/978-3-030-62529-0_10
- Cigna F, Tapete D (2021) Sentinel-1 big data processing with P-SBAS InSAR in the geohazards exploitation platform: an experiment on coastal land subsidence and landslides in Italy. *Remote Sensing* 13:1–26. <https://doi.org/10.3390/rs13050885>
- De Serio F, Armenio E, Mossa M, Petrillo AF (2018) How to define priorities in coastal vulnerability assessment. *Geosciences (Switzerland)* 8:1–20. <https://doi.org/10.3390/geosciences8110415>
- Debortoli NS, Clark DG, Ford JD et al (2019) An integrative climate change vulnerability index for Arctic aviation and marine transportation. *Nat Commun* 10:1–15. <https://doi.org/10.1038/s41467-019-10347-1>
- Demirkesen AC, Evrendilek F, Berberoglu S (2008) Quantifying coastal inundation vulnerability of Turkey to sea-level rise. *Environ Monit Assess* 138:101–106. <https://doi.org/10.1007/s10661-007-9746-7>
- Dereli MA, Tercan E (2020) Assessment of shoreline changes using historical satellite images and geospatial analysis along the Lake Salda in Turkey. *Earth Sci Inf* 13:709–718. <https://doi.org/10.1007/s12145-020-00460-x>
- Dierssen HM, Theberge AE (2016) Bathymetry: Assessment. *Encyclopedia of Natural Resources: Water* 629–636. <https://doi.org/10.1081/e-enrw-120048588>
- Done T, Jones R (2006) Tropical coastal ecosystems and climate change prediction: global and local risks. In: Done T (ed) *Coastal and Estuarine Studies*, 1st edn. American Geophysical Union, Melbourne, pp 5–32
- Dong L, Wang C, Tang Y et al (2021) Sentinel-1 InSAR Measurements of Land Subsidence over Henan Province, China from 2018 to 2020. 2021 SAR in Big Data Era, BIGSARDATA 2021 - Proceedings 1:2020–2023. <https://doi.org/10.1109/BIGSARDATA53212.2021.9574177>
- Doukakis E (2005) Coastal vulnerability and risk parameters. *European Water* 11:3–7
- El-Shahat S, El-Zafarany AM, El Seoud TA, Ghoniem SA (2021) Vulnerability assessment of African coasts to sea level rise using GIS and remote sensing. *Environ Dev Sustain* 23:2827–2845. <https://doi.org/10.1007/s10668-020-00639-8>
- Emran A (2016) Modeling spatio-temporal shoreline and areal dynamics of coastal island using geospatial technique. *Modeling Earth Systems and Environment* 1–11. <https://doi.org/10.1007/s40808-015-0060-z>
- Ermini L, Catani F, Casagli N (2005) Artificial neural networks applied to landslide susceptibility assessment. *Geomorphology* 66:327–343. <https://doi.org/10.1016/j.geomorph.2004.09.025>
- Fafalios S, Charonyktakis P, Tsamardinos I (2020) Gradient boosting trees. *Gnosis Data Analysis PC*. Heraklion, Greece, pp. 1–13. https://www.gnosisda.gr/wpcontent/uploads/2020/07/Gradient_Boosting_Implementation.pdf
- Firman T (2009) The continuity and change in mega-urbanization in Indonesia: a survey of Jakarta-Bandung Region (JBR) development. *Habitat Int* 33:327–339. <https://doi.org/10.1016/j.habitint.2008.08.005>
- Gayathri R, Bhaskaran PK, Jose F (2017) Coastal inundation research: An overview of the process. *Curr Sci* 112:267–278. <https://doi.org/10.18520/cs/v112/i02/267-278>
- Ghazali NHM, Awang NA, Mahmud M, Mokhtar A (2018) Impact of sea level rise and tsunami on coastal areas of North-West Peninsular Malaysia. *Irrig Drain* 67:119–129. <https://doi.org/10.1002/IRD.2244>
- Ghoussein Y, Mhawej M, Jaffal A et al (2018) Vulnerability assessment of the South-Lebanese coast: a GIS-based approach. *Ocean Coast Manag* 158:56–63. <https://doi.org/10.1016/J.OCECOAMAN.2018.03.028>
- Giannakidou C, Diakoulaki D, Memos CD (2020) Vulnerability to coastal flooding of industrial urban areas in Greece. *Environ Process* 7:749–766. <https://doi.org/10.1007/S40710-020-00442-7/TABLES/9>
- Gómez-Pazo A, Payo A, Paz-Delgado MV, Delgadillo-Calzadilla MA (2022) Open digital shoreline analysis system: ODSAS v1.0. *J Mar Sci Eng* 10:1–18. <https://doi.org/10.3390/jmse10010026>
- Gopinath S, Srinivasamoorthy K, Saravanan K et al (2016) Modeling saline water intrusion in Nagapattinam coastal aquifers, Tamilnadu, India. *Model Earth Syst Environ* 2:1–10. <https://doi.org/10.1007/s40808-015-0058-6>
- Gornitz V (1990) Vulnerability of the East Coast, U SA to Future Sea Level. *J Coast Res* 201–237
- Grandini M, Bagli E, Visani G (2020) Metrics for multi-class classification: an overview: arXiv. <https://doi.org/10.48550/arXiv.2008.05756>
- Haddad M, Hassani H, Taibi H (2013) Sea level in the Mediterranean Sea: seasonal adjustment and trend extraction within the framework of SSA. *Earth Sci Inf* 6:99–111. <https://doi.org/10.1007/s12145-013-0114-6>
- Hammid AT, Bin SMH, Abdalla AN (2018) Prediction of small hydro-power plant power production in Himreen Lake dam (HLD) using artificial neural network. *Alex Eng J* 57:211–221. <https://doi.org/10.1016/j.aej.2016.12.011>
- Handiani DN, Heriati A, Gunawan WA (2022) Comparison of coastal vulnerability assessment for subang regency in North Coast West Java-Indonesia. *Geomat Nat Haz Risk* 13:1178–1206. <https://doi.org/10.1080/19475705.2022.2066573>
- Hedger RD, Atkinson PM, Malthus TJ (2001) Optimizing sampling strategies for estimating mean water quality in lakes using geostatistical techniques with remote sensing. *Lakes Reserv Res Manag* 6:279–288. <https://doi.org/10.1046/j.1440-1770.2001.00159.x>
- Himmelstoss E, Henderson R, Kratzmann M, Farris A (2018) Digital Shoreline Analysis System (DSAS) Version 5.0 User Guide. Open-File Report 2018–1179 126
- Hossain SA, Mondal I, Thakur S, Fadhil Al-Quraishi AM (2022) Coastal vulnerability assessment of India's purba Medinipur-Balasore coastal stretch: a comparative study using empirical models. *Int J Disaster Risk Reduction* 77:103065. <https://doi.org/10.1016/j.ijdrr.2022.103065>
- Husnayaen RAB, Osawa T et al (2018) Physical assessment of coastal vulnerability under enhanced land subsidence in Semarang, Indonesia, using multi-sensor satellite data. *Adv Space Res* 61:2159–2179. <https://doi.org/10.1016/j.asr.2018.01.026>
- Huu NT, Huynh T (2018) Adaptation to salinity intrusion for rice farming household in the Vietnamese Mekong Delta. 6th Asian Academic Society International Conference (AASIC) 673–680
- Jafariroodsari S, Nourani V, Gökçekuş H (2021) Investigating sea-level change on the coastal aquifer, case study: Jafakendeh aquifer. *Model Earth Syst Environ* 7:2643–2651. <https://doi.org/10.1007/s40808-020-01043-x>
- Jamali A (2020) Land use land cover mapping using advanced machine learning classifiers: a case study of Shiraz city, Iran. *Earth Sci Inform* 13:1015–1030. <https://doi.org/10.1007/s12145-020-00475-4>
- Kabiri K (2017) Accuracy assessment of nearshore bathymetry information retrieved from Landsat-8 imagery. *Earth Sci Inf* 10:235–245. <https://doi.org/10.1007/s12145-017-0293-7>
- Kasim F, Siregar VP (2012) Coastal vulnerability assessment using integrated-method of CVI-MCA: a case study on the coastline of Indramayu. *Forum Geografi* 26(1):65–76

- Kathiravan K, Natesan U, Vishnunath R (2019) Developing GIS based coastal water quality index for Rameswaram Island, India positioned in Gulf of Mannar marine biosphere reserve. *Model Earth Syst Environ* 5:1519–1528. <https://doi.org/10.1007/s40808-019-00656-1>
- Kaur M, Sood SK (2020) Hydro-meteorological hazards and role of ICT during 2010–2019: a scientometric analysis. *Earth Sci Inf* 13:1201–1223. <https://doi.org/10.1007/s12145-020-00495-0>
- Ke G, Meng Q, Finley T et al (2017) LightGBM: A highly efficient gradient boosting decision tree. In: Proceedings: Adv Neural Inf Proces Syst, pp. 3147–3155
- Kementerian PPN/Bappenas (2020) Executive summary-climate resilience development policy 1st edn. In: Medrilzam Pratiwi S, Utomo, ET (eds.) Jakarta: ADB-JICAUSAID, pp. 2020–2045
- Khadijah, Saraswati R, Wibowo A (2020) Coastline changes on the Coast of Cirebon using landsat. E3S Web of Conferences 202:3–4. <https://doi.org/10.1051/e3sconf/202020215016>
- Krishnan P, Ananthan PS, Purvaja R et al (2019) Framework for mapping the drivers of coastal vulnerability and spatial decision making for climate-change adaptation: a case study from Maharashtra, India. *Ambio* 48:192–212. <https://doi.org/10.1007/s13280-018-1061-8>
- Krzywinski M, Altman N (2017) Points of significance classification and regression trees. *Nat Methods* 14:757–759
- Kumar BP, Babu KR, Anusha BN, Rajasekhar M (2022) Geo-environmental monitoring and assessment of land degradation and desertification in the semi-arid regions using Landsat 8 OLI / TIRS, LST, and NDVI approach. *Environ Challenges* 8. <https://doi.org/10.1016/j.envc.2022.100578>
- Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159. <https://doi.org/10.2307/2529310>
- Lewis RJ (2000) An introduction to classification and regression tree (CART) analysis. Presented at the 2000 Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California
- Li S, Xu W, Li Z (2022) Review of the SBAS InSAR time-series algorithms, applications, and challenges. *Geodesy Geodyn* 13:114–126. <https://doi.org/10.1016/j.geog.2021.09.007>
- Liang X, Guan Q, Clarke KC et al (2021) Understanding the drivers of sustainable land expansion using a patch-generating land use simulation (PLUS) model: A case study in Wuhan, China. *Comput Environ Urban Syst* 85. <https://doi.org/10.1016/j.compeurbysys.2020.101569>
- Lin W, Sun Y, Nijhuis S, Wang Z (2020) Scenario-based flood risk assessment for urbanizing deltas using future land-use simulation (FLUS): Guangzhou metropolitan area as a case study. *Sci Total Environ* 739:139899. <https://doi.org/10.1016/j.scitotenv.2020.139899>
- Loh W-Y, Eltinge J, Cho MJ, Li Y (2019) Classification and regression trees and forests for incomplete data from sample surveys. *Stat Sin* 29:431–453. <https://doi.org/10.5705/ss.202017.0225>
- Luu C, Bui QD, von Meding J (2021) Mapping direct flood impacts from a 2020 extreme flood event in Central Vietnam using spatial analysis techniques. *Int J Disaster Resilience Built Environ.* <https://doi.org/10.1108/IJDRBE-07-2021-0070>
- Mahamoud A, Gzam M, Ahmed Mohamed N et al (2022) A preliminary assessment of coastal vulnerability for Ngazidja Island, Comoros Archipelago, Western Indian Ocean. *Environ Earth Sci* 81:1–23. <https://doi.org/10.1007/s12665-021-10136-4>
- Malik A, Abdalla R (2016) Geospatial modeling of the impact of sea level rise on coastal communities: application of Richmond, British Columbia, Canada. *Model Earth Syst Environ* 2:1–17. <https://doi.org/10.1007/s40808-016-0199-2>
- Mandal S, Islam MS, Biswas MHA (2021) Modeling the potential impact of climate change on living beings near coastal areas. *Model Earth Syst Environ* 7:1783–1796. <https://doi.org/10.1007/s40808-020-00897-5>
- Mani Murali R, Ankita M, Amrita S, Vethamony P (2013) Coastal vulnerability assessment of Puducherry coast, India, using the analytical hierarchical process. *Nat Hazard* 13:3291–3311. <https://doi.org/10.5194/nhess-13-3291-2013>
- Manunta M, De Luca C, Zinno I et al (2019) The parallel SBAS approach for Sentinel-1 interferometric wide swath deformation time-series generation: algorithm description and products quality assessment. *IEEE Trans Geosci Remote Sens* 57:6229–6281. <https://doi.org/10.1109/TGRS.2019.2904912>
- Marfai MA (2011a) Impact of coastal inundation on ecology and agricultural land use case study in Central Java, Indonesia. *Quaectiones Geographicae* 30:19–32. <https://doi.org/10.2478/v10117-011-0024-y>
- Marfai MA (2011b) The hazards of coastal erosion in Central Java, Indonesia: an overview. *Geografia* 3:1–9
- Marfai MA, King L (2007) Monitoring land subsidence in Semarang, Indonesia. *Environ Geol* 53:651–659. <https://doi.org/10.1007/s00254-007-0680-3>
- Marfai MA, King L (2008) Potential vulnerability implications of coastal inundation due to sea level rise for the coastal zone of Semarang city, Indonesia. *Environ Geol* 54:1235–1245. <https://doi.org/10.1007/s00254-007-0906-4>
- Martínez Prentice R, Villoslada Peciña M, Ward RD et al (2021) Machine learning classification and accuracy assessment from high-resolution images of coastal wetlands. *Remote Sensing* 13:1–27. <https://doi.org/10.3390/rs13183669>
- McLaughlin S, Andrew J, Cooper G (2010) A multi-scale coastal vulnerability index: a tool for coastal managers? *Environ Hazards* 9:233–248. <https://doi.org/10.3763/EHAZ.2010.0052>
- Metelka V, Baratoux L, Jessell MW et al (2018) Automated regolith landform mapping using airborne geophysics and remote sensing data, Burkina Faso, West Africa. *Remote Sens Environ* 204:964–978. <https://doi.org/10.1016/j.rse.2017.08.004>
- Ministry of Maritime Affairs and Fisheries, Republic of Indonesia (2009) Coastal Vulnerability Index Map of Indonesia, Agency for Marine and Fisheries Research, accessed from <https://pusri-skdl.litbang.kkp.go.id/index.php/peta-kerentanan-pesisir-nasional#>. Accessed 17 Oct 2022
- Montesinos-López O, Montesinos A, Crossa J (2022) Overfitting, model tuning, and evaluation of prediction performance. https://doi.org/10.1007/978-3-030-89010-0_4
- Nasiri A, Shafiee N, Zandi R (2021) Spatial analysis of factors influencing land subsidence using the OLS model (Case study: Fahlian Aquifer). *Earth Sci Inf* 14:2133–2144. <https://doi.org/10.1007/s12145-021-00688-1>
- Natarajan L, Sivagnanam N, Usha T et al (2021) Shoreline changes over last five decades and predictions for 2030 and 2040: a case study from Cuddalore, southeast coast of India. *Earth Sci Inf* 14:1315–1325. <https://doi.org/10.1007/S12145-021-00668-5/FIGURES/8>
- Nath A, Koley B, Saraswati S, Ray BC (2021) Identification of the coastal hazard zone between the areas of Rasulpur and Subarnarekha estuary, east coast of India using multi-criteria evaluation method. *Model Earth Syst Environ* 7:2251–2265. <https://doi.org/10.1007/s40808-020-00986-5>
- Ndehedehe CE, Usman M, Okwuashi O, Ferreira VG (2022) Modeling impacts of climate change on coastal West African rainfall. *Model Earth Syst Environ* 8:3325–3340. <https://doi.org/10.1007/s40808-021-01302-5>
- Nguyen TTX (2015) Coastal vulnerability assessment: a case study in Kien Giang, western part of the Mekong River Delta in Vietnam. Thesis Collection 1954–2016 393
- Nikijuluw VPH (2017) Coastal Resources Conservation in Indonesia: Issues, Policies, and Future Directions. *Sumatra J Disaster Geogr Geogr Educ* 1:15. <https://doi.org/10.24036/sjdgge.v1i1.31>

- Nittrouer CA, Brunskill GJ, Figueiredo AG (1995) Importance of tropical coastal environments. *Geo-Mar Lett* 15:121–126. <https://doi.org/10.1007/BF01204452>
- Noi Phan T, Kuch V, Lehnert LW (2020) Land cover classification using google earth engine and random forest classifier-the role of image composition. *Remote Sensing* 12:1–22. <https://doi.org/10.3390/RS12152411>
- Noor NM, Abdul Maulud KN (2022) Coastal vulnerability: a brief review on integrated assessment in Southeast Asia. *J Mar Sci Eng* 10. <https://doi.org/10.3390/jmse10050595>
- Nurshodikin M, Saputra S (2021) Analisa Pemanfaatan Ruang Wilayah Pesisir Coastal Area Kabupaten Karimun. *Pelita Kota* 2:25–35
- Orieschnig CA, Belaud G, Venot JP et al (2021) Input imagery, classifiers, and cloud computing: insights from multi-temporal LULC mapping in the Cambodian Mekong Delta. *Eur J Remote Sens* 54:398–416. <https://doi.org/10.1080/22797254.2021.1948356>
- Özyurt G, Ergin A (2009) Application of sea level rise vulnerability assessment model to selected coastal areas of Turkey. *J Coastal Res* 2009:248–251
- Pandey A, Jain A (2017) Comparative analysis of KNN algorithm using various normalization techniques. *Int J Comput Netw Inf Secur* 9:36–42. <https://doi.org/10.5815/ijcnis.2017.11.04>
- Parthasarathy A, Natesan U (2015) Coastal vulnerability assessment: a case study on erosion and coastal change along Tuticorin, Gulf of Mannar. *Nat Hazards* 75:1713–1729. <https://doi.org/10.1007/s11069-014-1394-y>
- Pendleton EA, Theiler ER, Williams SJ (2005) Coastal Vulnerability Assessment of Gateway National Recreation Area (GATE) to Sea-Level Rise. Open-File Report 2004–1257; US Geological Survey 1–27
- Pham BT, Prakash I, Jaafari A, Bui DT (2018) Spatial prediction of rainfall-induced landslides using aggregating one-dependence estimators classifier. *J Indian Soc Remote Sens* 46:1457–1470. <https://doi.org/10.1007/s12524-018-0791-1>
- Pradeep J, Shaji E, Chandran CSS et al (2022) Assessment of coastal variations due to climate change using remote sensing and machine learning techniques: a case study from west coast of India. *Estuar Coast Shelf Sci* 275:107968. <https://doi.org/10.1016/j.ecss.2022.107968>
- Pramanik MK, Biswas SS, Mondal B, Pal R (2016) Coastal vulnerability assessment of the predicted sea level rise in the coastal zone of Krishna-Godavari delta region, Andhra Pradesh, east coast of India. *Environ Dev Sustain* 18:1635–1655. <https://doi.org/10.1007/s10668-015-9708-0>
- Pramova E, Chazarin F, Locatelli B, Hoppe M (2013) Climate Change Impact Chains in Tropical Coastal Areas. Bonn and Eschborn, Germany
- Prasetia R (2021) Coastal vulnerability analysis of the North Coast of Central Java Province. Delft Institute. MSc Thesis Identifier: WSE-CEPD, pp. 21–10. <https://doi.org/10.25831/bpr3-ds31>
- Prasetyo Y, Bashit N, Sasmito B, Setianingsih W (2019) Impact of land subsidence and sea level rise influence shoreline change in the Coastal Area of Demak. *IOP Conf Ser Earth Environ Sci* 280. <https://doi.org/10.1088/1755-1315/280/1/012006>
- Praticò S, Solano F, Di Fazio S, Modica G (2021) Machine learning classification of mediterranean forest habitats in google earth engine based on seasonal sentinel-2 time-series and input image composition optimisation. *Remote Sens* 13:1–28. <https://doi.org/10.3390/rs13040586>
- Purwaka TH, Sunoto (1997) Coastal resources management in Indonesia: legal and institutional aspects. Institutional issues and perspectives in the management of fisheries and Coastal Resources in Southeast Asia. ICLARM Technical Report 60 212
- Qi C, Fourie A, Chen Q, Zhang Q (2018) A strength prediction model using artificial intelligence for recycling waste tailings as cemented paste backfill. *J Clean Prod* 183:566–578. <https://doi.org/10.1016/j.jclepro.2018.02.154>
- Radiarta IN, Erlania, Haryadi J (2022) Coastal Vulnerability Assessment along the North Java Coastlines-Indonesia. *Jurnal Segara* 8:1–12
- Rao KN, Subraelu P, Rao TV et al (2009) Sea-level rise and coastal vulnerability: an assessment of Andhra Pradesh coast, India through remote sensing and GIS. *J Coast Conserv* 12:195–207. <https://doi.org/10.1007/s11852-009-0042-2>
- Remondo J, Gonzales-Diez A, Diaz de Teran JR, Cendrero A (2003) Landslide susceptibility models utilizing spatial data analysis technique. A case study from the lower Debra Valley, Guipúzcoa (Spain). *Nat Hazards* 30:267–279
- Ridwansyah I, Yulianti M, Apip et al (2020) The impact of land use and climate change on surface runoff and groundwater in Cimanuk watershed, Indonesia. *Limnology* 21:487–498. <https://doi.org/10.1007/s10201-020-00629-9>
- Rizzo A, Vandelli V, Buhagiar G et al (2020) Coastal vulnerability assessment along the north-eastern sector of Gozo Island (Malta, Mediterranean Sea). *Water (Switzerland)* 12. <https://doi.org/10.3390/w12051405>
- Roukounis CN, Tsirhrintzis VA (2022) Indices of coastal vulnerability to climate change: a review. *Environ Process* 9. <https://doi.org/10.1007/s40710-022-00577-9>
- Sahoo B, Bhaskaran PK (2018) Multi-hazard risk assessment of coastal vulnerability from tropical cyclones – a GIS based approach for the Odisha coast. *J Environ Manage* 206:1166–1178. <https://doi.org/10.1016/J.JENVMAN.2017.10.075>
- Samanta S, Paul SK (2016) Geospatial analysis of shoreline and land use/land cover changes through remote sensing and GIS techniques. *Model Earth Syst Environ* 2:1–8. <https://doi.org/10.1007/s40808-016-0180-0>
- Sankari TS, Chandramouli AR, Gokul K et al (2015) Coastal vulnerability mapping using geospatial Technologies in Cuddalore-Pichavaram Coastal Tract, Tamil Nadu, India. *Aquatic Procedia* 4:412–418. <https://doi.org/10.1016/j.aqpro.2015.02.055>
- Saikh AA, Pathan AI, Waikhom SI et al (2022) Application of latest HEC-RAS version 6 for 2D hydrodynamic modeling through GIS framework: a case study from coastal urban floodplain in India. *Model Earth Syst Environ*. <https://doi.org/10.1007/s40808-022-01567-4>
- Sidiq TP, Gumilar I, Meilano I et al (2021) Land subsidence of Java North Coast observed by SAR Interferometry. *IOP Conf Ser Earth Environ Sci* 873. <https://doi.org/10.1088/1755-1315/873/1/012078>
- Sim J, Wright CC (2005) The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 85:257–268. <https://doi.org/10.1093/ptj/85.3.257>
- Singh D, Singh B (2020) Investigating the impact of data normalization on classification performance. *Appl Soft Comput* 97:105524. <https://doi.org/10.1016/j.asoc.2019.105524>
- Sola J, Sevilla J (1997) Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Trans Nucl Sci* 44:1464–1468. <https://doi.org/10.1109/23.589532>
- Solihuddin T, Husrin S, Salim HL et al (2021) Coastal erosion on the north coast of Java: Adaptation strategies and coastal management. *IOP Conf Ser Earth Environ Sci* 777. <https://doi.org/10.1088/1755-1315/777/1/012035>
- Stammer D, Cazenave A, Ponte RM, Tamisiea ME (2012) Causes for contemporary regional sea level changes. <https://doi.org/10.1146/annurev-marine-121211-172406>
- Sudha NN, Satyanarayana RAN, Prasad V et al (2015) Coastal vulnerability assessment studies over India: a review. *77:405–428*. <https://doi.org/10.1007/s11069-015-1597-x>

- Suhardi I, Saraswati R, Abubakar R (2020) Perubahan Garis Pantai Pesisir Utara Jawa, 1st edn. Geography Department, Indonesia University, Jakarta, Indonesia
- Sui L, Wang J, Yang X, Wang Z (2020) Spatial-temporal characteristics of coastline changes in Indonesia from 1990 to 2018. *Sustainability* (Switzerland) 12:1–28. <https://doi.org/10.3390/SU12083242>
- Suwandana E (2019) Dinamika morfologi pantai Kabupaten Tangerang Banten dan Pantai Indah Kapuk Jakarta melalui analisis citra google earth. *Jurnal Perikanan Dan Kelautan* 9:55–68
- Taalab K, Cheng T, Zhang Y (2018) Mapping landslide susceptibility and types using random Forest. *Big Earth Data* 2:159–178. <https://doi.org/10.1080/20964471.2018.1472392>
- Talukdar S, Singha P, Mahato S, Pal S (2020) Land-use land-cover classification by machine learning classifiers for satellite observations—A review. *Remote Sensing* 12:24. <https://doi.org/10.3390/rs12071135>
- Tanim AH, McRae CB, Tavakol-davani H, Goharian E (2022) Flood Detection in urban areas using satellite imagery and machine learning. *Water* (Switzerland) 14. <https://doi.org/10.3390/w14071140>
- Tellman B, Sullivan JA, Kuhn C, Kettner AJ, Doyle CS, Brakenridge GR, Erickson T, Slayback DA (2021) Satellites observe increasing proportion of population exposed to floods. *Nature*. <https://doi.org/10.1038/s41586-021-03695-w>
- Thanh Noi P, Kappas M (2018) Comparison of random forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors* 18:18. <https://doi.org/10.3390/s18010018>
- Thieler ER, Hammar-Klose ES (1999) National assessment of coastal vulnerability to sea level rise: preliminary results for the U.S. Atlanta Coast USGS, Open File Report, pp. 99–593. <http://pubs.usgs.gov/of/1999/of99-593/index.html>
- Tian S, Zhang X, Tian J, Sun Q (2016) Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China. *Remote Sens* 8:1–14. <https://doi.org/10.3390/rs8110954>
- Tran KT, Nguyen HDQ, Truong PT et al (2022) Evaluation of sea-level rise influence on tidal characteristics using a numerical model approach: a case study of a southern city coastal area in Vietnam. *Model Earth Syst Environ*. <https://doi.org/10.1007/s40808-022-01536-x>
- Triana K, Wahyudi AJ (2020) Sea level rise in Indonesia: The drivers and the combined impacts from land subsidence. *ASEAN J Sci Technol Dev* 37:115–121. <https://doi.org/10.29037/AJSTD.627>
- Val AL, De Almeida-Val VMF, Randall DJ (2006) Tropical environment. *Physiol Trop Fish* 21:1–45. [https://doi.org/10.1016/S1546-5098\(05\)21001-4](https://doi.org/10.1016/S1546-5098(05)21001-4)
- Warnadi, A'Rachman FR, Hijrawadi SN (2020) Spatiotemporal shoreline change analysis in the downstream area of Cisadane Watershed since 1972. *IOP Conf Ser Earth Environ Sci* 412. <https://doi.org/10.1088/1755-1315/412/1/012007>
- Wibowo PLA, Hartoko A, Ambariyanto A (2015) Land Subsidence Affects Coastal Zone Vulnerability (Pengaruh Penurunan Tanah Terhadap Kerentanan Wilayah Pesisir). *ILMU KELAUTAN: Indonesian J Mar Sci* 20:127. <https://doi.org/10.14710/ik.ijms.20.3.127-134>
- World Bank and Asian Development Bank (2021) Climate Risk Country Profile-Indonesia, World Bank Group
- Wu T (2021) Quantifying coastal flood vulnerability for climate adaptation policy using principal component analysis. *Ecol Ind* 129:108006. <https://doi.org/10.1016/j.ecolind.2021.108006>
- Wu Q, Yue H, Liu Y, Hou E (2022) Geospatial quantitative analysis of the Aral Sea shoreline changes using RS and GIS techniques. *Earth Sci Inf* 15:137–149. <https://doi.org/10.1007/s12145-021-00714-2>
- Yankey RK, Anornu GK, Appiah-Adjei EK et al (2020) Structural equation modeling and GIS application into non-carcinogenic health risk assessment of the phreatic aquifers of the southwestern coastal basin-Ghana. *Model Earth Syst Environ* 6:2553–2564. <https://doi.org/10.1007/s40808-020-00851-5>
- Yingli LV, Le QT, Bui HB et al (2020) A comparative study of different machine learning algorithms in predicting the content of ilmenite in titanium placer. *Appl Sci* (Switzerland) 10. <https://doi.org/10.3390/app10020635>
- Yuwono BD, Abidin HZ, Gumilar I et al (2016) Preliminary survey and performance of land subsidence in North Semarang Demak. *AIP Conference Proceedings* 1730. <https://doi.org/10.1063/1.4947410>
- Zanaga D, Van De Kerchove R, De Keersmaecker W, Souverijns N, Brockmann C, Quast R, Wevers J, Grosu A, Paccini A, Vergnaud S, Cartus O, Santoro M, Fritz S, Georgieva I, Lesiv M, Carter S, Herold M, Li, Linlin, Tsednbazar NE, Ramoino F, Arino O (2021). *ESA WorldCover 10 m 2020 v100*. 10.5281/zenodo.5571936
- Zhao Y, Zhu W, Wei P et al (2022) Classification of Zambian grasslands using random forest feature importance selection during the optimal phenological period. *Ecol Ind* 135:108529. <https://doi.org/10.1016/j.ecolind.2021.108529>
- Zheng Z, Zha H, Zhang T et al (2008) A general boosting method and its application to learning ranking functions for Web search. In: *Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference*. Sunnyvale, pp 1–8
- Theobald D, Harrison-Atlas D, Monahan W (2015) Ecologically-relevant maps of landforms and physiographic diversity for climate adaptation planning. *PLoS ONE* 10(12). <https://doi.org/10.1371/journal.pone.0143619>
- Muraina IO (2022) Ideal dataset splitting ratios in machine learning algorithms: general concerns for data scientists and data analysts. *7th International Mardin Artuklu Scientific Research Conference 2022*. https://www.researchgate.net/publication/35824895_IDEAL_DATASET_SPLITTING RATIOS_IN_MACHINE LEARNING_ALGORITHMS_GENERAL CONCERNS FOR DATA SCIENTISTS
- Nguyen QH, Ly HB, Ho LS, Al-Ansari N, Le HV, Tran VQ, Prakash I, Pham BT (2021) Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering* 2021:15. <https://doi.org/10.1155/2021/483284>
- Birba D (2020) A Comparative study of data splitting algorithms for machine learning model selection, Dissertation, 2020. KTH ROYAL INSTITUTE OF TECHNOLOGY ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
- Cohen JA (1960) Coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1(20):37–46. <https://doi.org/10.1177/00131644600200104>
- Montesinos López OA, Montesinos López A, Crossa J (2022) Random forest for genomic prediction. In: *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer, Cham. https://doi.org/10.1007/978-3-030-89010-0_15
- Ao Y, Li H, Zhu L, Ali S, Yang Z (2018) Logging lithology discrimination in the prototype similarity space with random forest - IEEE Geoscience and Remote Sensing Letters. <https://doi.org/10.1109/LGRS.2018.2882123>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.