JSCDM

JOURNAL OF SOFT COMPUTING AND DATA MINING
e-ISSN: 2716-621X

Vol. 7 No. (2026) 349-370
https://publisher.uthm.edu.my/ojs/index.php/jscdm

Optimized Tsunami Vulnerability Area Classification Using Hybrid Fuzzy-SVM Model Through Spatial Data Processing

Adi Susilo[1,2], Renaldi Primaswara Prasetya[3,4], Muhammad Fathur Rouf Hasan[2,3]*, Rizal Setya Perdana[5], Azizul Azhar Ramli[6], Rony Prianto Nugraha[7]

[1] *Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Malang, 65145, INDONESIA*
[2] *Center Study on Geosciences and Hazard Mitigation, Universitas Brawijaya, Malang, 65145, INDONESIA*
[3] *Graduate School, Universitas Brawijaya, Malang, 65145, INDONESIA*
[4] *Department of Electrical Engineering and Informatics, Faculty of Engineering Universitas Negeri Malang, Malang, 65145, INDONESIA*
[5] *Department of Computer Science, Faculty of Computer Science, Universitas Brawijaya, Malang, 65145, INDONESIA*
[6] *Faculty of Computer Sciences and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400, Johor, MALAYSIA*
[7] *Department of Engineering Science, The University of Auckland, Private Bag, Auckland 90210, NEW ZEALAND*

*Corresponding Author: rouf@ub.ac.id
DOI: https://doi.org/10.30880/jscdm.2026.07.02.019

Article Info

Received: 8 November 2025
Accepted: 7 March 2026
Available online: 30 June 2026

Keywords

Classification, fuzzy-SVM, hybrid model, Tsunami, vulnerability, spatial data

Abstract

Tsunami disasters pose serious threats to human life and coastal infrastructure and require accurate mapping of tsunami-prone areas for effective disaster mitigation and coastal planning. Machine learning methods, including weighted overlay and Support Vector Machine (SVM), are widely used but often struggle to represent gradual transitions between vulnerability classes. This study proposes a hybrid fuzzy–SVM approach to enhance the accuracy and robustness of tsunami vulnerability classification. Three geospatial parameters, elevation, land cover, and inundation extent, were used as primary inputs, each transformed through fuzzy membership functions to handle uncertainty and spatial ambiguity. The fuzzy-transformed variables were aggregated into a normalized Fuzzy Vulnerability Index (FVI), which was subsequently classified using SVM with linear and RBF kernels under a one-vs-rest scheme to generate vulnerability maps for the southern coast of East Java. Experimental results demonstrated that the proposed hybrid fuzzy–SVM outperformed both conventional SVM and weighted overlay methods. The model achieved an overall accuracy of 91.3%, precision of 0.911, recall of 0.910, and F1-score of 0.910, indicating strong agreement between predicted and reference vulnerability maps. Overall, the hybrid fuzzy–SVM framework provides a more flexible and data-driven approach to tsunami vulnerability assessment.

1. Introduction

Tsunamis rank among the most destructive natural hazards along coastlines worldwide, inflicting heavy losses on lives and infrastructure [1]. In Indonesia, the threat is especially acute: the archipelago straddles the Pacific "Ring of Fire," and much of Java and southern Sumatra lie above the Sunda subduction zone, where the Indo-Australian Plate dives beneath Eurasia [2]. These geological conditions cause frequent strong tsunami waves, necessitating effective tsunami risk mitigation efforts in Indonesia's coastal areas. Therefore, mapping and classifying tsunami-prone areas can be a first step in reducing the risk.

Therefore, many studies use GIS-based multi-criteria methods, such as weighted overlay methods, to map tsunami-vulnerable zones. For example, Isdianto et al. [3] combined factors such as terrain slope, elevation, land use, and distance from the coast, each reclassified and weighted to produce a tsunami-susceptibility map. This study used the weighted overlay method to combine various vulnerability factors, including land slope, elevation, land use, and distance from the coast. The same scale was applied to each parameter, and the weighted values were summed. These weighted sums were then categorized into vulnerability levels.

Current research integrates artificial intelligence to improve vulnerability mapping models. Bibliometric studies show an upward trend in the use of geographic information systems (GIS) techniques and machine learning for tsunami prediction. For instance, Hou et al. report that machine learning classifiers, such as Random Forests, are widely used for satellite-based land-cover classification in tsunami risk studies [4]. In Indonesia, Wibowo et al. [5] developed a tsunami vulnerability model using Landsat 8 data. They extracted vegetation indices, including NDVI and NDWI, and trained KNN, Random Forest, and SVM classifiers to predict vulnerability levels. These data-driven models can capture complex spatial patterns of risk that simpler overlay methods may miss.

However, weighted overlay and SVM methods have limitations. Weighted-overlay maps produce crisp risk categories, so areas falling near the boundary between classes can be misclassified [6-7]. For example, Guntur et al. define five vulnerability classes from very low to very high; any gradual gradient within a zone is flattened into a single category [8]. Likewise, a standard SVM classifier will impose a sharp decision boundary even where vulnerability changes smoothly. Moreover, these methods often rely on fixed inputs, such as static terrain layers, historical earthquake indicators, and expert-derived weights, which may not fully capture evolving or uncertain hazard scenarios.

To address transitional and ambiguous zones between vulnerability classes, we implemented a hybrid fuzzy-SVM framework. Fuzzy logic allows partial class membership, smoothing the transitions between risk levels. In multi-criteria GIS, fuzzy membership functions with values 0–1 are frequently used to account for uncertainty. For example, Ghadamode et al. [9] transformed weighted criteria into fuzzy membership scores, deriving tsunami vulnerability classes that closely matched observed inundation patterns. Their fuzzy-based model yielded a graded risk map from very low to very high that agreed well with modelled tsunami heights.

Despite the growing body of research on tsunami vulnerability assessment, several limitations remain. First, many studies rely on deterministic weighted overlay methods that apply fixed thresholds and lack uncertainty modeling. Second, recent machine learning approaches often prioritize predictive accuracy over interpretability in spatial vulnerability classification. Third, although fuzzy logic has been introduced to address gradual class transitions, existing implementations present important constraints. For instance, GIS-based fuzzy–AHP approaches remain highly dependent on expert-defined weights, potentially introducing subjectivity and limiting reproducibility. Similarly, rule-based fuzzy inference systems often focus on hazard alert classification rather than spatial vulnerability mapping and lack empirical validation against standardized ground truth datasets. Therefore, this study aims to bridge these gaps by proposing a fuzzy-SVM-based vulnerability framework that emphasizes gradual classification, interpretability, and empirical validation against official risk data.

The hybrid framework is designed to overcome the inherent limitation of conventional crisp classification, which fails in properly representing transitional zones between "moderately vulnerable" and "highly vulnerable" areas. The proposed model combines fuzzy logic's ability to handle uncertainty and gradual class boundaries with the robustness of SVMs to construct optimal separating hyperplanes. This study makes a key contribution by developing a hybrid fuzzy-SVM framework for tsunami vulnerability mapping that improves accuracy in handling transition zones and ambiguity between vulnerability classes. Furthermore, this research employs a fine-scale multiclass mapping method in the southern coastal region of East Java, Indonesia, yielding a more adaptive and precise spatial representation for disaster mitigation planning.

2. Related Work

Previous studies have developed various approaches using technologies such as GIS and remote sensing to obtain spatial data. These studies have even collaborated with AI-based methods or machine learning to determine and classify areas with tsunami potential. In this section, we will review related research on the same case study using the SVM method and other machine learning approaches. We will also discuss the limitations of each previous study. In connection with tsunami vulnerability classification using the SVM machine learning method, SVM has demonstrated robustness to high-dimensional data, making it well-suited to the complex, multi-dimensional nature of the tsunami prediction dataset. Tsunami prediction often relies on a wide range of variables, including seismic data, oceanographic parameters, and historical tsunami records, which together span a high-dimensional feature space.

Research conducted by Sukmana [10] compared SVM and Random Forest using earthquake features. The study classified tsunami-prone areas using a secondary data approach with historical earthquake data. Although the data spanned a long period from 2001 to 2023, this approach did not account for spatial factors, even though real-time data are crucial for determining tsunami vulnerability classification [11]. The resulting accuracy was 65.61%, with a precision of 70.59% and a recall of 19.67%. Prasetyo and Maulana's subsequent study used SVMs on seismic data—specifically earthquake depth and magnitude—to predict tsunami potential in Indonesia [12]. This approach achieved 97% accuracy using a polynomial kernel. However, depth and magnitude do not directly affect tsunami vulnerability. A series of other factors technically affect tsunami magnitude, thereby influencing the determination of areas potentially vulnerable to tsunamis [13].

Additionally, the generated class output is limited to two classes: vulnerable and non-vulnerable. Thus, the classification type is simple, even though this study tested four SVM kernel types. Wibowo et al. conducted a subsequent study that mapped tsunami-prone areas in the Banyuwangi region using an SVM approach combined with the k-Nearest Neighbors (k-NN) method [5]. However, this study did not use a Digital Elevation Model (DEM) or an inundation model, as did Isnaeny's studies [14]. It focused solely on vegetation indices, though determining tsunami-prone areas is influenced by multiple factors, including DEM and inundation [15-16], and even the vulnerability map consisted of only three classes.

Mardiani [17] evaluated several classification methods, namely k-Nearest Neighbors (k-NN), Naïve Bayes, Decision Tree, and ensemble methods (Random Forest, AdaBoost) to classify tsunami occurrence from earthquakes. The k-NN model achieved the highest AUC of 94.4% and a precision of 82.8%. However, the data used was limited to historical earthquake data, and no spatial analysis was considered. Siregar et al. conducted the same research [18], using historical earthquake features based on magnitude measurements and an optimized neural network with SMOTE to predict tsunamis from large-earthquake data with magnitudes greater than 5. Their model achieved 99.43% accuracy and an F1-score of 97.9%.

Further research by Sambah et al. mapped tsunami-prone areas in East Java, specifically in Jember. His two studies used parameters such as slope, elevation, inundation, land cover, and distance from the coast and land. He used the Weighted Overlay method to account for all parameters [19-20]. However, this method still relied on manual expert judgment. The determination of weights was subjective, as was the case with mapping on Bali Island [21]. This study lacked comparative data and validation of the mapping results. The study focused solely on producing potential area maps, without considering the accuracy of the resulting data.

Recent studies have demonstrated the effectiveness of fuzzy logic in tsunami vulnerability assessment, given its ability to model uncertainty and gradual class transitions. Ghadamode et al. [9] integrated GIS-based fuzzy membership functions with AHP weighting to generate multiclass vulnerability maps. However, their approach relies heavily on expert-driven weights, which may introduce subjectivity. Similarly, the study by Rehman & Zhang [22] applied fuzzy inference to tsunami hazard zonation but focused primarily on rule-based classification, without machine-learning validation.

Table 1 *Summary of related papers*

Author	Research Area	Classification Method	Input Parameters	Output	Limitations
Isnaeni & Prasetyo [14]	Bantul, Indonesia - Subdistrict level	ANN, SVM, RF	NDVI, NDBI, NDWI, SAVI, MNDWI	Risk of tsunami land damage	Not integrated with the DEM or the spatial modelling of inundation. The focus is only on the vegetation index.
Wibowo et al. [5]	Banyuwangi, Indonesia - Subdistrict level	SVM, RF, k-NN	NDVI, NDWI, MSAVI, MNDWI (Landsat-8)	Tsunami vulnerability map (high, medium, low)	Not integrated with the DEM or the spatial modelling of inundation. The focus is only on the vegetation index.
Mardiani et al. [17]	Indonesia	KNN, Random Forest, Naïve Bayes, Decision Tree	Depth and Magnitude	Generates two categories: “Vulnerable” and “Not Vulnerable.”	Historical earthquake data, including depth and magnitude, is the only thing to focus on; do not use DEM or inundation models.
Sukmana et al. [10]	Indonesia	SVM, Random Forest	Earthquake features	Generates two categories: “Vulnerable” and “Not Vulnerable.”	No spatial analysis. Focus only on historical earthquake data.
Siregar et al. [18]	Indonesia	SMOTE	Earthquake features	Generates two categories: “Vulnerable” and “Not Vulnerable.”	No spatial analysis. Focus only on historical earthquake data.
Prasetyo & Maulana. [12]	Indonesia	SVM	Depth and Magnitude	Generates two categories: “Vulnerable” and “Not Vulnerable.”	Uses depth data and magnitude. Meanwhile, classification is limited to two classes only.
Sambah et al. [19][20]	Jember, East Java, Indonesia - Subdistrict level	Weighted Overlay	Elevation, Land cover, inundation	Tsunami vulnerability map (low, slightly low, medium, slightly high, high)	Each feature is classified separately, without being combined, and the classification is based on subjective expert judgment.
Sambah et al. [21]	Bali, Indonesia - Subdistrict level	Weighted Overlay	Elevation, Land cover, Inundation, Run up	Tsunami vulnerability map (low, slightly low, medium, slightly high, high)	Each feature is classified separately, without being combined, and the classification is based on subjective expert judgment.
Ghadamode et al. [9]	Andaman Region, Bay of Bengal	Fuzzy, AHP	Elevation, slope, shoreline distance, vegetation density, run up	Tsunami vulnerability map (low, slightly low, medium, slightly high, high)	Approach relies heavily on expert-driven weights, which may introduce subjectivity
Rehman & Zhang [22]	Makran, Coastal zone	Fuzzy	Earthquake features, Downgoing slab velocity	Tsunami alert system	Focus on alert prediction systems, not explicit spatial vulnerability mapping.

Based on the studies presented in Table 1, although weighted overlay [19-21] and geographic information system-based index methods [5][14] are widely used due to their simplicity and interpretability, they inherently rely on crisp thresholding and fixed weight assignments. Such deterministic structures may fail to capture gradual transitions in vulnerability levels, particularly in coastal regions with high spatial heterogeneity. This limitation motivates the exploration of more flexible modeling frameworks that can handle uncertainty and fuzzy class boundaries.

Recent studies (2022–2025) have increasingly applied machine learning techniques, such as SVMs, Random Forests, and deep learning models, to improve classification accuracy [10], [12], [17-18]. While these approaches demonstrate strong predictive performance, most of them treat vulnerability classes as strictly separable categories and rely on hard class boundaries. Furthermore, uncertainty propagation and transitional spatial zones are rarely modeled explicitly. As a result, the interpretability and reliability of classification outputs in borderline areas remain limited. Although fuzzy logic has been employed in several disaster risk studies to address uncertainty [9], [22], many implementations remain limited to standalone fuzzy overlay schemes without integration into supervised classification frameworks. The combination of fuzzification for uncertainty modeling with machine learning classifiers for decision boundary optimization remains underexplored in tsunami vulnerability assessment, particularly in Southeast Asian coastal contexts.

The existing approaches to tsunami vulnerability mapping can be broadly categorized into deterministic geographic information system-based overlays, machine learning classifiers, and standalone fuzzy logic models. Deterministic methods suffer from rigid thresholding, while conventional machine learning models often neglect uncertainty modeling in transitional zones. Meanwhile, fuzzy-based methods rarely integrate supervised classification for optimizing decision boundaries. Therefore, there remains a need for an integrated framework that combines uncertainty-aware fuzzification with data-driven classification to produce both accurate and interpretable multiclass vulnerability maps. This study addresses this gap by proposing a hybrid fuzzy–SVM framework tailored for multiclass tsunami vulnerability assessment.

3. Methodology

This study employs a hybrid Fuzzy–Support Vector Machine (SVM) approach to map tsunami vulnerability levels at the sub-district scale along the southern coast of East Java Province. The hybrid framework is designed to overcome the inherent limitation of conventional crisp classification. The methodological workflow comprises several stages.

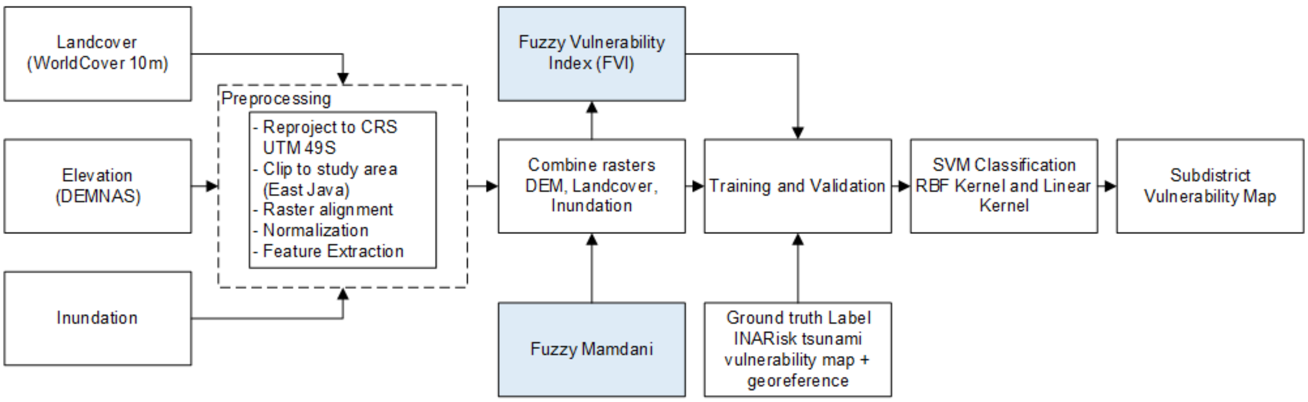


Fig. 1 *Tsunami vulnerability area framework*

First, geospatial preprocessing was conducted by harmonizing coordinate reference systems (CRS), clipping the study area using administrative boundaries, and generating thematic layers including: (1) elevation from the *DEMNAS* dataset; (2) landcover from the *ESA WorldCover 2020* dataset; and (3) tsunami inundation extent, derived from both elevation-based bathtub modelling and simulated run-up scenarios. All thematic layers were aggregated to the sub-district level using *zonal statistics* in QGIS to ensure spatial consistency among datasets. In this hybrid implementation, fuzzy logic is used for pre-classification, converting input indices such as elevation, land cover, and inundation into membership degrees that account for spatial ambiguity. The defuzzification result, FVI, becomes an additional feature that makes the support vector machine (SVM) more sensitive to transition zones.

The multiclass classification scheme uses one-vs-rest (OvR) by combining SVM probabilities and fuzzy membership, which can be used in the final decision stage to reduce errors in highly ambiguous areas. This approach produces more detailed vulnerability maps and provides a data-driven classification framework that can be replicated and validated with official national-scale risk data such as InaRISK. In short, this study's major contribution is the presentation of a hybrid fuzzy-SVM method combining two kernels to classify potential tsunami-vulnerability areas. Furthermore, the study provides comprehensive physical modelling and simulation results in the form of zone maps down to the subdistrict level that are potentially vulnerable to tsunamis. Figure 1 visualizes the framework of the proposed research study.

3.1 Study Area

The study area covers the entire East Java region (Figure 2), extending from Pacitan to Banyuwangi and bordering the Indian Ocean. The geology of this region is influenced by the subduction zone along the southern coast of Java Island [23]. This tectonic interaction makes East Java's southern coastline highly vulnerable to earthquakes and tsunamis. For instance, Pacitan Regency is considered one of the areas with the highest tsunami risk along Java's southern coast. Geographically, the area consists of a narrow lowland on the coast backed by Karst mountains. Several strong tectonic earthquakes have been recorded in this region in recent years. For example, an intraplate earthquake with a magnitude of 6.1 occurred south of Malang on April 10, 2021, killing at least 10 people and damaging more than 3,000 buildings [24]. An in-depth analysis shows that this earthquake occurred at a depth of about 60 km within the subduction zone, so the potential for a tsunami is relatively small.

Additionally, a magnitude 6.0 earthquake with a similar mechanism was recorded south of Malang on May 15, 2021. No major tsunamis have been recorded off the southern coast of East Java in the past 5–6 years. However, the region's history of strong earthquakes and the condition of the southern Java fault segment, which has the potential to generate a megathrust tsunami, highlight the importance of mitigation, particularly at the sub-district level, to enable more specific and targeted disaster risk mapping.

Fig. 2 *The research area is located in East Java Province*

3.2 Dataset

This study uses three main multi-criteria datasets to map tsunami potential on the southern coast of East Java, as follows:

- Land Cover: The 2020 ESA WorldCover global land cover map with a spatial resolution of 10 meters was used. This dataset is based on Sentinel-1 and Sentinel-2 imagery and covers 11 land cover classes, including forest cover, built-up land, agricultural land, and water. ESA WorldCover's advantages are its high resolution (10 meters) and up-to-date global coverage. This enables it to capture land-use details in coastal areas effectively. According to the official WorldCover data source, the 2020 WorldCover product was independently validated with an overall accuracy of 74–75%, making it reliable for spatial vulnerability analyses at the district level. WorldCover data is available in GeoTIFF (COG) format for each 3×3-degree tile and can be processed in QGIS for land classification.

- Digital Elevation Model: Indonesia's official elevation data from BIG (DEMNAS) is used to model land topography. DEMNAS is the result of combining IFSAR, TerraSAR-X, ALOS PALSAR satellite imagery, and other stereo mass distribution data. DEMNAS products have a spatial resolution of 0.27 arc seconds, equivalent to 8.25 meters, and a vertical datum of EGM2008 [25]. With a resolution of approximately 8 meters, DEMNAS can display contour details and slope gradients in coastal and hilly areas of East Java more accurately than ASTER or SRTM. DEMNAS is available in a standardized grid format covering the entire territory of Indonesia, facilitating tsunami inundation analysis.
- Inundation: Several studies have demonstrated the influence of tsunami magnitude, height, and run-up on inundation through simulation and empirical formula analyses. One approach is to simulate tsunami inundation using numerical models for a specific earthquake scenario. For instance, one study simulated tsunami wave propagation using the Shallow Water Equations Planar Wave Equations with hydrodynamic software such as Delft3D, adapting a hypothetical megathrust earthquake scenario with a magnitude of 8.8 in South Java [23]. The simulation models the earthquake source, seafloor deformation, and wave propagation to the coast. The simulation results provide maps of maximum wave heights along the coastline and of maximum land inundation. These maps are classified as components of the tsunami hazard. The output is typically a water depth map for each earthquake scenario. The second approach is an empirical estimation model-based tsunami inundation simulation. This simulation uses empirical or parametric models to estimate wave heights and the extent of tsunami inundation based on earthquake and coastal-slope characteristics. One common method is the tsunami wave height loss (H-loss) equation by Berryman [26], which relates wave attenuation to distance travelled, coastal slope, and surface roughness. This formula estimates the initial run-up height along the coast and decreases it exponentially with increasing distance from the coast. Another parametric relationship method, such as that proposed by Abe (1993), provides an empirical equation for tsunami run-up (Ht) that demonstrates a logarithmic relationship between magnitude (Mw) and tsunami height (Ht), as illustrated by Eq. 1.

$$log_{10}H_t = M_w - log_{10}R - 5.55 + C \tag{1}$$

C equals 0 for the fore-arc and 1 for the back-arc, and R represents the horizontal distance from the epicentre to the coast at which the wave makes landfall [27]. The southern coast of East Java lies in the fore-arc zone (C=0) because it is closer to the tsunami source zone (the megathrust) and lies in front of the volcanic arc. The output from this empirical model is run-up values, or maximum water heights, at a given distance from the coast. These values can be used to estimate an inundation area based on terrain. This approach is useful for this study because R represents the distance that the tsunami propagates from the earthquake's epicentre to the coast, or the length of the path in the sea affected by the earthquake's magnitude. Thus, the value of R indirectly determines the run-up and the inundation distance from the coast, allowing us to determine how far inland the tsunami can reach (inundation) using Eq. 2.

$$D_{inundation} = \frac{H_t}{tan\,\theta} \tag{2}$$

Where D is the distance from the inundation to the land, and θ is the angle of the land slope with the stipulation that the coastline has the gentlest slope.

3.3 Landcover Vulnerability Index (LC-Index)

The Land Cover (LC) feature uses the WorldCover dataset, which provides very high-level detail. The data will be processed to identify land-use features, including settlements, rice fields, mangrove forests, and bodies of water. This is based on the assumption that land use type affects an area's vulnerability to tsunamis. The main stages for obtaining the land cover vulnerability index value are described below.

- **Stage 1 -** Data preparation and preprocessing.
  Land cover information was derived from the ESA WorldCover 2020 dataset, which has a spatial resolution of 10 meters. This dataset was clipped to cover the southern coastal area of East Java Province. After clipping the dataset to the study area, the projection system was adjusted to be consistent with other spatial data.

- **Stage 2 -** Reclassification based on vulnerability level
  Each landcover class was assigned a relative vulnerability score on a scale of 0 (least vulnerable) to 1 (most vulnerable), based on its potential to increase or mitigate tsunami damage. Built-up areas and bare coastal lands received the highest scores, while forested or vegetated regions received the lowest. This scoring was adapted from previous tsunami-exposure studies in Southeast Asia and calibrated using expert judgment.

- **Stage 3 -** Calculating composite indices per subdistrict
  The proportion of land cover area in each class is calculated for each subdistrict relative to its total area. The land cover vulnerability index (LC index) is calculated using the following formula:

$$LC\_index_{kec} = \sum_{j=1}^{m} (\frac{B_j}{A_{tot}}) \ x \ W_j \tag{3}$$

  where B_j = and cover the area of class j, A_{tot} = subdistrict area, W_j = and vulnerability weight. Using the same steps as in the elevation calculation, the proportion of land cover in each subdistrict determines the land cover type in that subdistrict, such as residential areas, rice fields, mangrove forests, or waterways.

- **Stage 4 –** Normalization
  The index results are normalized to the range [0, 1]. Values close to 1 indicate areas with land cover types that are highly vulnerable to tsunami inundation.

3.4 Elevation Vulnerability Index (Elev-Index)

The preprocessing stages for elevation data begin with converting the spatial mapping data to UTM 49, which produces a more accurate area, and clipping the data to the East Java region. After stacking the spatial elevation data with the spatial data of East Java, raster alignment arranges the raster layers to match each other in the same space. This ensures that pixels from different layers can be correctly connected. The Coordinate Reference System (CRS) determines the exact size and area of the cells that will be adjusted to the aligned raster layers. The ultimate goal is to perform zonal statistics by converting binary areas to raster layers to categorize parts of a subdistrict based on weighted proportions or elevation indices. The sequential steps for obtaining the elevation vulnerability index are described in the subsequent stages.

- **Stage 1 -** Digital Elevation Model preparation
  Elevation data were derived from the *National DEM (DEMNAS)* with a spatial resolution of approximately 8–30 meters. The DEM was mosaicked and clipped to the study area, ensuring topographic consistency and alignment with the coastal boundary. The dataset represents terrain elevation relative to Mean Sea Level (MSL).

- **Stage 2 -** Derivation of elevation statistics
  For each sub-district, the average elevation and its distribution were extracted. This stage aimed to characterize the relative elevations of coastal areas, with lower elevations corresponding to greater susceptibility to tsunami inundation.

- **Stage 3 -** Classification of elevation into vulnerability classes
  Elevation values were grouped into categorical vulnerability classes based on empirical thresholds derived from previous tsunami studies, e.g., <5 m, 5–10 m, 10–20 m, >20 m. Each class was assigned a vulnerability score ranging from 0 to 1, with lower terrain receiving higher scores. Table 2 shows the weighting score parameters for each class.

The weighting score parameters for each input class used to determine the vulnerability index for each criterion are shown in the following table.

Table 2 *Parameter classification based on vulnerability classes*

Vulnerability Class	Elevation (m)	Landuse	Run-up
5	<5	Urban	>16
4	5-10	Agriculture	*6-16*
3	10-15	Bare soil	*2-6*
2	15-20	Water	0.75-2
1	>20	Forest	< 0.75

- **Stage 4 -** Index calculation per subdistrict
  The proportion of area in each elevation class was computed and multiplied by its respective score. The aggregated *Elevation Vulnerability Index* per sub-district was formulated as:

$$Elev\_Index_{kec} = \sum_{i=1}^{n} \left(\frac{A_i}{A_{tot}}\right) \ x \ W_i \tag{4}$$

  where A_i is the area of elevation class-i in the subdistrict, A_{tot} is the total area of the subdistrict, dan W_i is the vulnerability weight for class-i. For example, the total area of Tambaksari Subdistrict is 10,000,000 m^2 or 10 km^2. After processing the data, the area of each raster in the subdistrict is determined based on how much of it falls into each of the five elevation classes. The proportion of each class is then calculated by dividing the area of each class by the total area of the subdistrict. Then, the proportion of each elevation class is multiplied by the predetermined weight and summed to determine the dominant value or elevation vulnerability score within the subdistrict.

- **Stage 5** – Normalization
  The index results were normalized so that their values fell within the range of [0,1]. Values close to 1 indicate areas that are highly vulnerable to tsunami inundation.

3.5 Inundation Vulnerability Index (INUND-Index)

The inundation index reflects two main aspects: the proportion of flooded areas and the extent of flooding inland. Additionally, the model will include two sources of inundation: simulation scenarios and empirical models.

- **Stage 1 -** Preparation of inundation scenarios
  As previously mentioned, we modelled the extent of the inundation using two complementary approaches:
  - A simulation-based scenario using numerical data from research conducted by Abe, based on equations 1 and 2. In this study, the 9.0 magnitude earthquake was used for the inundation scenario. This is based on the fact that the largest earthquake ever recorded in East Java was 9.0 on the Richter scale and has never exceeded 9.0 over the last 50 years of earthquake history. The earthquake is located 230 km from the coastline, in the center of the megathrust area. This illustrates a large earthquake, which is then used as a reference value in an empirical approach scenario.
  - An empirical model based on the bathtub approach, in which all cells below a specified run-up height threshold and connected to the coastline were classified as inundated. In this study, the run-up and inundation values will be obtained using Eqs. 1 and 2, which have been presented previously and will be presented in the next stage.

- **Stage 2 -** Determination of run-up parameters
  The run-up scenario described in Stage 1 will determine the landward extent of inundation for each subdistrict. The tsunami wave height used for the run-up value is 12.26 m, calculated using Eq. 1. $\log_{10} H_t = 9 - \log_{10} 230 - 5.55 + 0 = 9 - 2.362 - 5.55 = 1.088$, which means the value of $Ht = 10^{1.088} = 12.26\ m$, resulting in a flood distance to land of 12.26 m / 0.005 = 2456 m or 2.456 km, in terms that the coastline has the gentlest slope. These results align with the data from studies [28], [29]. For instance, in Aceh, a substantial run-up of 30 to 34 meters was recorded on the west coast, inundating the mainland for approximately 5 kilometers. In Sulawesi, the 2018 Palu earthquake showed a run-up of around 10-11 m at several points, resulting in flooding up to 488 m in the Layana area. In Tohoku, Japan, a run-up of 39.7 m caused flooding over more than 5 km.

- **Stage 3 -** Calculating the inundation index
  Vulnerability to inundation is determined by the percentage of flooded area, calculated by dividing each subdistrict's flooded area by its total area.

3.6 Fuzzification

The fuzzy methods are optimized and combined with the SVM before classification. After obtaining index values for each input variable, ranging from the elevation index and land cover index to the inundation index, the values are converted into fuzzy membership degrees (μ) based on the triangular function. This stage is important because fuzzy logic begins to refine the boundaries between classes, thereby reducing the uncertainty around them. For instance, the distinction between areas that are slightly high or slightly low and those that are medium is no longer biased. This is achieved through the triangular membership function (trimf) with parameters a, b, and c ($a \leq b \leq c$), using the following formula:

$$\mu(x; a, b, c) = \begin{cases} 0, \mathrm{x} \leq a \\ \dfrac{x-a}{b-a}, a < x \leq b \\ \dfrac{c-x}{c-b}, b < x < c \\ 0, \mathrm{x} \geq \mathrm{c} \end{cases} \tag{5}$$

Three fuzzy sets were defined for each index: low, medium, and high elevation for the elevation index; vegetated (low vulnerability), mixed (medium vulnerability), and built-up (high vulnerability) for the land cover index; and low, medium, and high inundation for the inundation index. These are all clarified and presented in Tables 3, 4, and 5.

Table 3 *Fuzzy membership degree for the elevation index*

Fuzzy set	Membership Function (μ)	Parameters (a,b,c)	Description
Low	trimf(x; 0.0, 0.2, 0.4)	0.0, 0.2, 0.4	Low elevation index means high vulnerability
Medium	trimf(x; 0.3, 0.5, 0.7)	0.3, 0.5, 0.7	Transition to the medium elevation zone
High	trimf(x; 0.6, 0.8, 1.0)	0.6, 0.8, 1.0	A high elevation index means low vulnerability

Table 4 *Fuzzy membership degree for the landcover index*

Fuzzy set	Membership Function (μ)	Parameters (a,b,c)	Description
Vegetated	trimf(x; 0.0, 0.2, 0.4)	0.0, 0.2, 0.4	Dominant vegetation areas
Mixed	trimf(x; 0.3, 0.5, 0.7)	0.3, 0.5, 0.7	Transitional areas, such as suburban and open land
Built-up	trimf(x; 0.6, 0.8, 1.0)	0.6, 0.8, 1.0	Densely populated areas

Table 5 *Fuzzy membership degree for the inundation index*

Fuzzy set	Membership Function (μ)	Parameters (a,b,c)	Description
Low	trimf(x; 0.0, 0.2, 0.4)	0.0, 0.2, 0.4	Minimal flooding area
Medium	trimf(x; 0.3, 0.5, 0.7)	0.3, 0.5, 0.7	Moderate flooding area
High	trimf(x; 0.6, 0.8, 1.0)	0.6, 0.8, 1.0	Extensive flooding area

Triangular membership functions were selected due to their interpretability, computational simplicity, and widespread use in spatial vulnerability modeling. The parameters (a, b, c) for each indicator were not arbitrarily assigned; instead, they were determined based on the normalized distribution of each variable (0–1 scale) and informed by established domain knowledge regarding tsunami exposure characteristics [19-21], [23].

Specifically, breakpoint values were aligned with quantile-based distribution patterns to ensure balanced representation of low, medium, and high vulnerability conditions. This data-driven calibration prevents extreme skewness while preserving the ordinal relationship among vulnerability levels. The use of consistent triangular structures across the three indices (elevation, land cover, and inundation extent) ensures comparability and stability during aggregation into the FVI.

Moreover, since the final FVI is normalized to the same 0–1 scale as the InaRISK index, the fuzzification design maintains structural consistency between the input transformation and the subsequent classification thresholds. This alignment enhances interpretability and methodological coherence across the modeling framework. The next step is to use Fuzzy Rule Inference to combine all these membership degrees and produce a final vulnerability score (FVI). This FVI value will later be used as an additional feature in the SVM input.

3.7 Fuzzy Rule Inference

The fuzzy inference process integrates various environmental conditions into a single, uncertainty-based decision-making framework [30]. Using the Mamdani model, the system assesses combinations of elevation, inundation, and land cover adaptively to generate confidence levels for each tsunami vulnerability category. The results of this inference are fuzzy membership degrees (μ_D) for each vulnerability class. These degrees are then converted into crisp values through a defuzzification stage. In this study, the inference system employs IF–THEN rules, wherein each rule maps parameter conditions to vulnerability levels in the form of general rules *IF (Elevation IS A) AND (Landcover IS B) AND (Inundation IS C) THEN (Vulnerability IS D)*, with A ∈ {Low, Medium, High}; B ∈ {Vegetated, Mixed, Built-up}; C ∈ {Low, Medium, High}; D ∈ {Low, Slightly Low, Medium, Slightly High, High}.

Table 6 *Fuzzy rule-based model of tsunami vulnerability*

No.	IF-THEN Rules
R1	IF (Elevation IS High) AND (Landcover IS Vegetated) AND (Inundation IS Low) THEN (Vulnerability IS Low)
R2	IF (Elevation IS High) AND (Landcover IS Mixed) AND (Inundation IS Medium) THEN (Vulnerability IS Slightly Low)
R3	IF (Elevation IS Medium) AND (Landcover IS Vegetated) AND (Inundation IS Low) THEN (Vulnerability IS Slightly Low)
R4	IF (Elevation IS Medium) AND (Landcover IS Mixed) AND (Inundation IS Medium) THEN (Vulnerability IS Medium)
….	…………
R15	IF (Elevation IS Low) AND (Landcover IS Built-up) AND (Inundation IS Medium) THEN (Vulnerability IS High)

The fuzzy rule base uses the Mamdani AND logic operator, which returns the minimum value from each antecedent (condition) and its THEN implication, as shown in Table 6 for all combinations of membership values, with results in Eq. 4.

$$\mu R_j = (\mu_{E,A}, \mu_{L,B}, \mu_{I,C}) \tag{6}$$

Where μR_j is the activation strength of rule-j, $\mu_{E,A}$ is the elevation membership degree for label A, $\mu_{L,B}$ is the land cover membership degree for label B, and $\mu_{I,C}$ is the inundation membership degree for label C. Then, the fuzzy membership degrees are generated through generalization in Eq. 5.

$$\mu D(x) = (max_j(min_i(\mu_{x_i}, A_{ij})) \tag{7}$$

Where x_i is the input variable (elevation, land cover, or inundation), A_{ij} is the fuzzy input for rule j, and D is the fuzzy label for the vulnerability output.

3.8 Defuzzification

Defuzzification was applied to convert fuzzy inference outputs into a single continuous vulnerability score per spatial unit. We used a weighted-average singleton approach, assigning representative numeric values to the five vulnerability classes as {0.00, 0.25, 0.50, 0.75, 1.00}. The FVI for each unit was computed as the normalized weighted sum of the class memberships, as shown in Eq. 6.

$$FVI = \frac{\sum_{k=1}^{K} \mu_k z_k}{\sum_{k=1}^{K} \mu_k} \tag{8}$$

Where K = 5, corresponding to the number of output classes, ranges from low to high, μ_k is the membership degree to which the output belongs to class k, resulting from the aggregation of inferences (value ∈ [0,1]), and the representative (singleton) numerical value for k class on a scale of 0 to 1. The resulting FVI (0–1) was appended to the feature vector (elev_index, lc_index, and inund_index) and standardized before SVM.

3.9 Vulnerability Classification Using SVM

This study uses a machine learning algorithm based on a support vector machine (SVM) with fuzzy output to classify tsunami vulnerability levels at the subdistrict level along the southern coast of East Java. Three main spatial variables are used as features: elevation index, land cover, inundation index, and additional FVI values. Each sample initially corresponds to a raster pixel extracted from the aligned geospatial layers. For model training, pixel-level features were used to preserve spatial detail and variability. Each variable is normalized using min-max scaling, resulting in values on a 0-1 scale.

The ground truth vulnerability data were obtained from the official InaRISK dataset provided by the National Disaster Management Agency (BNPB). The dataset was accessed in August, 2025, representing the latest available national tsunami vulnerability zoning at the time of analysis. InaRISK provides a continuous vulnerability index ranging from 0 to 1, which is categorized into three primary classes: Low (0.00–0.30), Medium (0.31–0.60), and High (0.61–1.00). To enable finer discrimination of transitional vulnerability zones, the continuous InaRISK index was reclassified into five levels in this study. The reclassification was performed through proportional subdivision of the original index intervals while preserving the conceptual boundaries defined by InaRISK. Specifically, the Low and Medium ranges were divided into two sub-intervals each, resulting in five final classes: Low (0.00–0.15), Slightly Low (0.16–0.30), Medium (0.31–0.45), Slightly High (0.46–0.60), and High (0.61–1.00).

This approach ensures consistency with the official InaRISK framework while allowing for more granular classification to better capture gradual transitions in vulnerability levels. The full class-mapping scheme is presented in Table 7. Evaluation was conducted at the subdistrict (polygon) level to ensure consistency with the administrative unit used in the InaRISK reference data. Specifically, pixel-level predictions within each subdistrict were aggregated using a majority-voting scheme to determine the representative vulnerability class of the polygon. The aggregated class was then compared against the official InaRISK vulnerability label assigned to that subdistrict. This two-stage design preserves fine-grained feature learning while ensuring that the final evaluation reflects administratively meaningful spatial units.

Table 7 *Class mapping from InaRISK index to five-level vulnerability scheme*

InaRISK Index Range	Original InaRISK Class	Reclassified Range	Final Class
0.00 – 0.30	Low	0.00 – 0.15	Low
		0.16 – 0.30	Slightly Low
0.31 – 0.60	Medium	0.31 – 0.45	Medium
		0.46 – 0.60	Slightly High
0.61 – 1.00	High	0.61 – 1.00	High

During training, 663 random samples were evenly distributed across classes. Then, the data were split into a 70% training set and a 30% test set using stratified sampling. Next, the hybrid fuzzy-SVM model was trained using linear and radial basis function (RBF) kernels, and the main parameters (C and γ) were optimized through several scenarios. Finally, the method was evaluated using a confusion matrix, as well as accuracy, precision, recall, and the F1 score.

3.9.1 Normalization

Generally, SVM calculations begin with data normalization. In this approach, normalization is necessary to equalize the data scale within a specified range, so that each attribute contributes equally to the model [31]. In SVM, normalization is important because many algorithms are sensitive to differences in feature scales. For instance, if one feature has a value range of 1–1,000 and another feature has a range of only 0–1, the feature with the larger range could dominate the training process. This means that the SVM will learn more from the dominant feature, and the contributions of the other features will be minimal, thereby disproportionately affecting the hyperplane. SVM normalization is carried out using Equation 7.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (9)$$

J. of Soft Computing and Data Mining Vol. 7 No. 2 (2026) p. 349-370

Before classification, all input variables, including elevation index, land cover index, inundation index, and Fuzzy Vulnerability Index (FVI), were standardized using min-max normalization to eliminate scale bias across heterogeneous parameters. The SVM model was then trained using both Linear and Radial Basis Function (RBF) kernels under a One-vs-Rest multiclass scheme.

3.9.2 Multiclass Classification

This study views tsunami vulnerability classification not as a binary problem (vulnerable or not vulnerable) but as having five main levels: low, slightly low, medium, slightly high, and high. To accommodate multiclass classification needs, the One-vs-Rest (OvR) strategy is employed. With the OvR approach, the system builds k classification models, where k is the number of classes. In this case (k = 5), five support vector machine (SVM) models are formed:

- Model 1: Low vs Others
  All samples labelled "low" were treated as the positive class (+1), while those labelled "slightly low," "medium," "slightly high," and "high" were grouped into the negative class (-1). The purpose of this model is to identify areas that are truly at low risk of tsunami impact. Its purpose is to distinguish zones that are truly safe from those that are potentially low-risk.
- Model 2: Slightly Low vs Others
  All samples labelled "slightly low" were treated as positive class (+1), while those labelled "low," "medium," "slightly high," and "high" were grouped into the negative class (-1). This model identifies areas with slightly below-average characteristics, such as high elevation, dominant vegetation cover, and low tsunami inundation.
- Model 3: Medium vs Others
  All samples labelled "medium" were treated as the positive class (+1), while those labelled "low," "slightly low," "slightly high," and "high" were grouped into the negative class (-1). This model identifies transition zones, such as areas with elevations between 5 and 15 meters, areas with settlements near the coast, and areas prone to medium tsunami inundation.
- Model 4: Slightly high vs. others
  All samples labelled "slightly high" were treated as positive class (+1), while those labelled "low," "slightly low," "medium," and "high" were grouped into the negative class (-1). This model identifies areas with high potential impact. These areas are priority zones for mitigation and typically cover densely populated coastal settlements or estuary plains.
- Model 5: High vs Others
  All samples labelled "high" were treated as the positive class (+1), while all other labels were combined into the negative class (–1). This model emphasizes identifying areas with extreme risk and maximum vulnerability. These areas include coastal regions with elevations below 5 meters, high population densities, and deep, extensive tsunami inundation zones.

During the prediction process, the five models were used to evaluate the spatial units in each subdistrict in East Java. Each model produced a decision score reflecting the level of confidence that the subdistrict belonged to the tested class. The class with the highest decision function score was selected as the final classification result.

3.9.3 SVM Kernel

Initially, the SVM was a linear classifier that classifies data using a straight separating line. SVM was then developed to solve non-linear problems by mapping data x into the function ϕ(x), which transforms the data into a higher-dimensional space where it can be linearly separated. SVM uses the kernel trick, a mathematical function that maps data from a low-dimensional space to a higher-dimensional space without explicitly computing coordinates in the new space [32]. Thus, data that was previously not linearly separable can be more easily separated.

Generally, the kernel function defines the degree of similarity between two data samples. Given two feature vectors, x_i and x_j, kernel $K(x_i x_j)$ produces a value indicating the degree of similarity between them. This kernel value is then used in the SVM optimization formulation to form a hyperplane that separates classes [33]. The various kernel functions used in this study are shown in Table 8. The linear and radial basis function (RBF) kernel values will later be entered into the decision function, $f(x)$ to determine the classification class.

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b \tag{10}$$

Where x is the feature vector of the test sample, x_i is the feature vector of the training sample used in SVM training, y_i is the class label of the x_i training sample, α_i is the Lagrange coefficient resulting from optimization in training, $K(x_i x_j)$ is the kernel function that calculates the similarity between the test sample and the training sample, b is the bias or intercept of the hyperplane, and $f(x)$ is the score of the decision function. In the Linear and Radial Basis Function (RBF) kernel, x_i and x_j denote input feature vectors in the transformed feature space, while $\|x_i - x_j\|^2$ represents the squared Euclidean distance between them. The parameter γ controls the width of the kernel and determines the influence range of each training sample in shaping the decision boundary.

Table 8 *Kernel list*

Kernel name	Kernel function
Linear Kernel	$K(x_i, x_j) = x_i . x_j$
Radial Basis Function (RBF)	$K(x_i, x_j) = exp\ (-\gamma \|x_i - x_j\|^2)$

3.10 Evaluation Measurement

Model performance is evaluated using a confusion matrix metric on test data, which measures classification model performance in machine learning. The matrix compares the model's predictions with the test data's actual labels to produce values for accuracy, precision, recall, and the F1 score. The matrix consists of four main components: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). In the case of a tsunami, TP is defined as when the actual condition is positive in the INARisk data or reference set, based on the relevant research subdistrict sample, and is predicted to be positive by hybrid Fuzzy-SVM. This means that both have the same true value. FP is data that should be negative in the INARisk data or reference set based on the relevant research subdistrict sample, but which is predicted correctly by the system. TN is negative data predicted correctly by the model, which is also negative. FN is data that is positive, but predicted as negative by the system. In the multi-class setting, these metrics are computed under a one-vs-rest scheme, where each class is treated as the positive class against all remaining classes.

Accuracy measures the total correctness of the model's classification by comparing the proportion of correctly classified areas to the total area tested. In the context of tsunami vulnerability mapping, accuracy reflects how consistent the hybrid model's classification results are with tsunami reference maps, such as observational data from previous research or simulations from official agencies. High accuracy indicates that the model's results match actual conditions.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

Precision is a measure of how accurately a model predicts a particular class. It is defined as the proportion of the area correctly predicted to belong to that class, compared to the total area predicted to belong to that class. In the context of tsunami vulnerability mapping, precision indicates how reliably the model identifies areas categorized as "tsunami-prone" without misclassifying safe areas. A high-precision score means the model rarely issues false alarms, so the mapping results are highly reliable.

$$precision = \frac{TP}{TP + FP} \tag{12}$$

Recall, also known as Sensitivity or True Positive Rate, measures the extent to which the model can detect all areas that truly belong to a particular vulnerability class. It is important to demonstrate a model's ability to capture all areas potentially affected by a tsunami. A high recall indicates that the model has successfully identified most high-risk areas, thereby minimizing the risk of information loss (omission error) in those areas. Conversely, a low recall indicates that the model has failed to detect tsunami-prone areas.

$$recall = \frac{TP}{TP + FN} \tag{13}$$

The F1 score is a measure of balance between precision and recall that provides a comprehensive evaluation of the model's performance. It is defined as the harmonic mean of precision and recall, as shown in Eq. 10. The F1-score is used to assess a model's stability between its ability to detect vulnerable areas and its accuracy in identifying them. A high F1 score indicates that the model has an optimal balance: it accurately detects high-risk areas without producing too many false predictions.

$$F1 - score = \ 2\,x\,\frac{precision * recall}{precision + recall} \quad (14)$$

4. Result and Discussion

Evaluation process of the ability of a hybrid fuzzy-SVM model to classify areas potentially vulnerable to tsunamis, we tuned the hyperparameters for each kernel. This allowed us to determine and assess the robustness of our predictions of potential tsunami-vulnerability areas at the subdistrict level. The trial involved maximizing the model via hyperparameter tuning, assessing model performance using a confusion matrix, and comparing results across different kernels. Additionally, tsunami-vulnerability areas were modeled at the sub-district level using QGIS tools.

4.1 Kernel SVM Hyperparameter Testing

The hybrid Fuzzy-SVM kernel was tested with two kernels: linear and RBF. Six possible SVM kernel parameters with various values were tested. Three configurations were evaluated for each kernel. In the linear kernel, the main parameter is C, which controls the level of regularization, with grid values of {0.1, 1, 10}. In the RBF kernel, two parameters were varied: C, with values {1, 10, 100}, and γ (gamma), with values {0.01, 0.1, 1}. Hyperparameter combinations were tested using a 5-fold cross-validation grid search (Table 9). The main evaluation metric was the F1-score, which balances precision and recall in cases of class imbalance in tsunami disasters.

The evaluation results show that the RBF kernel performs better than the linear kernel on average. The optimal model was achieved using C = 10 and γ = 1, yielding an accuracy of 91.3% and an F1-score of 0.91. These results demonstrate the model's exceptional ability to distinguish between tsunami-prone and non-prone areas, highlighting the RBF kernel's superior capacity to capture intricate patterns. Meanwhile, the linear kernel achieved a maximum accuracy of 85.2% and an F1-score of 0.85. Although these results are quite good, the linear kernel is less capable of capturing the non-linear complexity in the relationships among the criteria, namely DEM, land cover, and inundation.

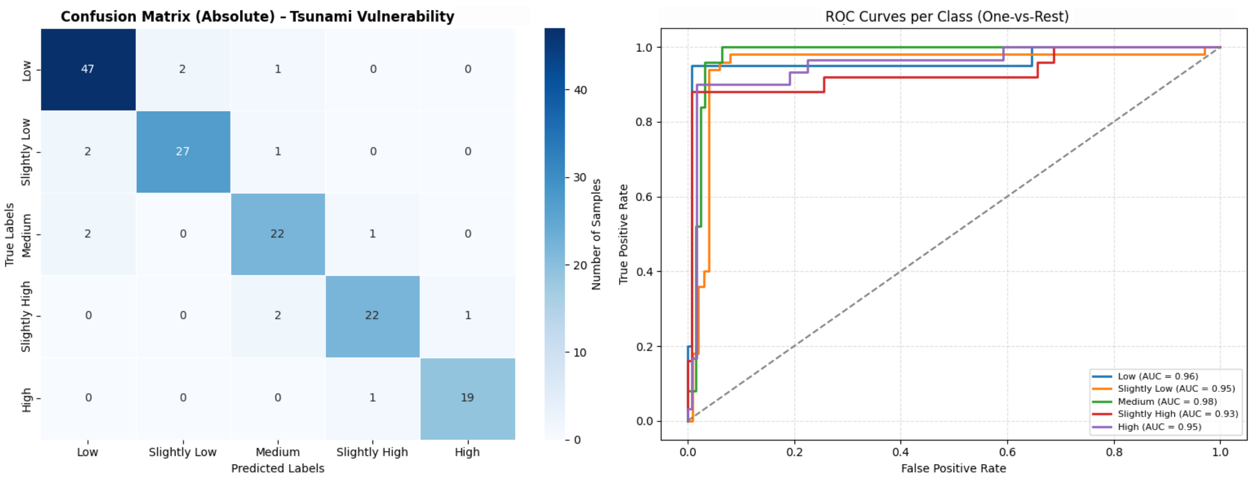
Table 9 *Kernel model performance evaluation*

Kernel	Coefficient C	Coefficient γ	Accuracy (%)	Precision	Recall	F1 Score
Linear	0.1	-	84.7	0.83	0.85	0.84
Linear	1	-	85.1	0.84	0.86	0.85
Linear	10	-	85.2	0.84	0.86	0.85
RBF	0.1	0. 01	87.9	0.86	0.89	0.87
RBF	1	0.1	89.4	0.88	0.90	0.89
RBF	10	1	91.3	0.91	0.91	0.91

The evaluation table results also showed that accuracy would increase with an increase in the C coefficient. This suggests that increasing the C value helps the model reduce the decision margin, allowing classes to be separated more effectively. Next, the gamma value controls the influence of the distance between points. A small gamma value widens the kernel function or smooths the boundary, while a large gamma value narrows the kernel, thereby increasing the risk of overfitting. Therefore, a gamma value that is too large can make the model overly sensitive to the training data. In this test, combining a larger C value with a balanced gamma value provided optimal results, separating classes more accurately without losing the model's ability to generalize to new data.

4.2 Performance Result

After determining the optimal parameter combination for each kernel, the next step is to assess the model's performance. Figure 3 shows the results of testing the RBF kernel with the highest accuracy, presented in the confusion matrix and ROC curve. At this stage, of the total 663 spatial sample points, 513 were allocated to training and 150 to testing. The test data were selected from 10 priority areas identified as vulnerable both historically and in literature sources, specifically the southern coastal subdistricts covering Pacitan [21], Trenggalek [34], Blitar [35], Malang [24], Lumajang [36], Jember [19], [20], Banyuwangi [5], Probolinggo, Tulungagung, and Situbondo. Each city has 15 subdistricts, for a total of 150 in the test data. These will also be compared with national INARisk data.

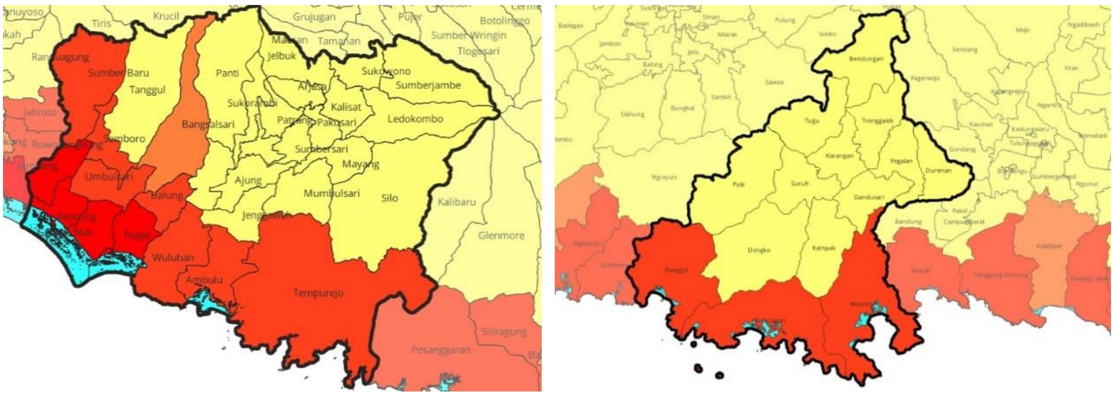


(a) **(b)**

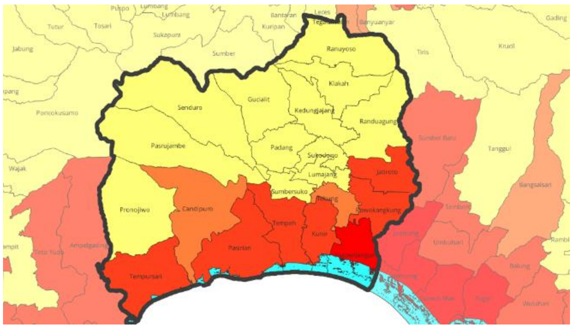
Fig. 3 *(a) Confusion matrix; (b) ROC curve of the proposed model for five classes of Tsunami vulnerability*

The performance of the hybrid fuzzy-SVM -based tsunami vulnerability classification was first evaluated using a confusion matrix to assess class-wise prediction accuracy. The confusion matrix is organized in ordinal class order: Low, Slightly Low, Medium, Slightly High, and High, reflecting increasing tsunami risk. This ordering ensures consistent interpretation of misclassification patterns, particularly between adjacent vulnerability levels. As shown in Figure 3(a), the model successfully identified a high proportion of true positives (TP) for all five vulnerability classes, particularly for the *low* and *high* categories. The correctly classified samples (TP) accounted for 47 out of 50 instances in the *low* class, 27 out of 30 for slightly *low*, 22 out of 25 for *medium*, 22 out of 25 for slightly *high*, and 19 out of 20 for *high* vulnerability zones, such as the Pacitan region, where the Pringkuku, Sudimoro, and Kebonagung subdistricts were found to have the same actual values.

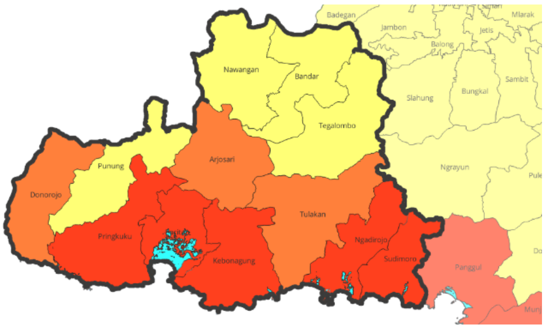
Another example is the Lumajang region, which successfully identified a high-category tsunami-prone area in the Yosowilangun subdistrict, as shown in Figure 4. A relatively small number of false positives (FP) and false negatives (FN) were recorded, mostly between adjacent classes such as *medium-slightly high* and slightly *low–medium* transitions. This pattern indicates that classification ambiguity occurs mainly in transitional coastal areas with steeper elevation gradients and greater land cover heterogeneity.



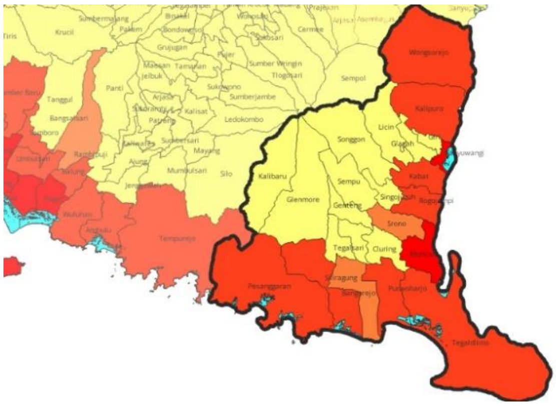
(a) **(b)**



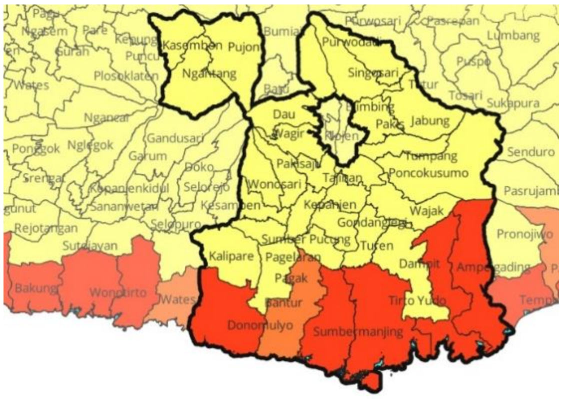
(c)



(d)



(e)



(f)

Fig. 4 *The following are examples of mapping data from hybrid model testing at the subdistrict level: (a) Subdistricts in the Jember region; (b) Subdistricts in the Trenggalek region; (c) Subdistricts in the Lumajang region; (d) Subdistricts in the Pacitan region; (e) Subdistricts in the Banyuwangi region; and (f) Subdistricts in the Malang region*

Further evaluation process of the model's discriminatory capabilities, a Receiver Operating Characteristic (ROC) analysis was performed under the One-vs-Rest (OvR) configuration, where each class is compared against all other classes combined, as shown in Figure 3(b). The model outputs were normalized to probability scores using a min-max transformation before calculating AUC. The reported AUC values represent macro-averaged performance across all classes. The resulting ROC curves show that all classes exhibit strong deviation from the random baseline, indicating robust separation between vulnerable and non-vulnerable zones. The computed Area Under the Curve (AUC) values ranged from 0.93 to 0.98, with the highest AUC observed for the Low and Moderate vulnerability classes, at 0.96 and 0.98, respectively. The relatively high AUC values across all categories confirm that the proposed model can effectively discriminate coastal areas characterized by low elevation (<10 m), dense built-up land cover, and wide inundation zones. The overall macro-average AUC of 0.95 further substantiates the model's ability to handle multiclass tsunami vulnerability mapping.

Overall, the hybrid fuzzy-SVM model with RBF kernel achieved an accuracy of 91.3% (Table 10). The macro-average precision value of 0.911 and the macro-average recall value of 0.910 demonstrate the balance between the model's accuracy and completeness in detecting each category of tsunami vulnerability. The highest precision and recall values appear in the "low" and "high" classes, at 0.94 and 0.95, respectively. This demonstrates the model's ability to identify completely safe areas and areas with extreme vulnerability levels without producing many false positives, resulting in a low error rate.

Table 10 *Test evaluation results for each class*

Class	Precision	Recall	F1-Score
Low	0.940	0.940	0.940
Slightly low	0.900	0.900	0.900
Medium	0.846	0.880	0.863
Slightly high	0.917	0.880	0.898
High	0.950	0.950	0.950
Macro average	**0.911**	**0.910**	**0.910**
Overall Accuracy			**0.913**

The relatively lower performance of the "Medium Vulnerability" class, with an F1 score of 0.863, suggests slight overlap in geospatial characteristics between medium- and high-risk areas. This overlap has no significant effect, especially in areas of coastal topography transition. However, the average macro F1 score of 0.910 indicates that the model is stable in detecting spatial variation across vulnerability classes. The hybrid fuzzy-SVM model using the radial basis function (RBF) kernel achieves high classification accuracy, indicating that a machine learning–based approach can capture the non-linear relationships among digital elevation model variables, land cover, and inundation, which affect the level of tsunami vulnerability in the southern coastal region of East Java.

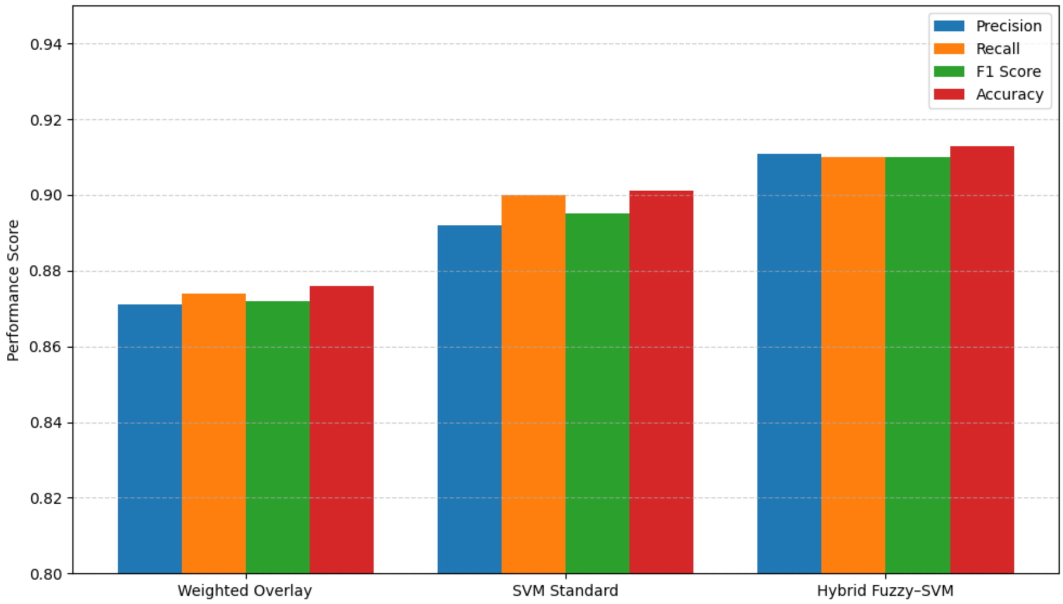
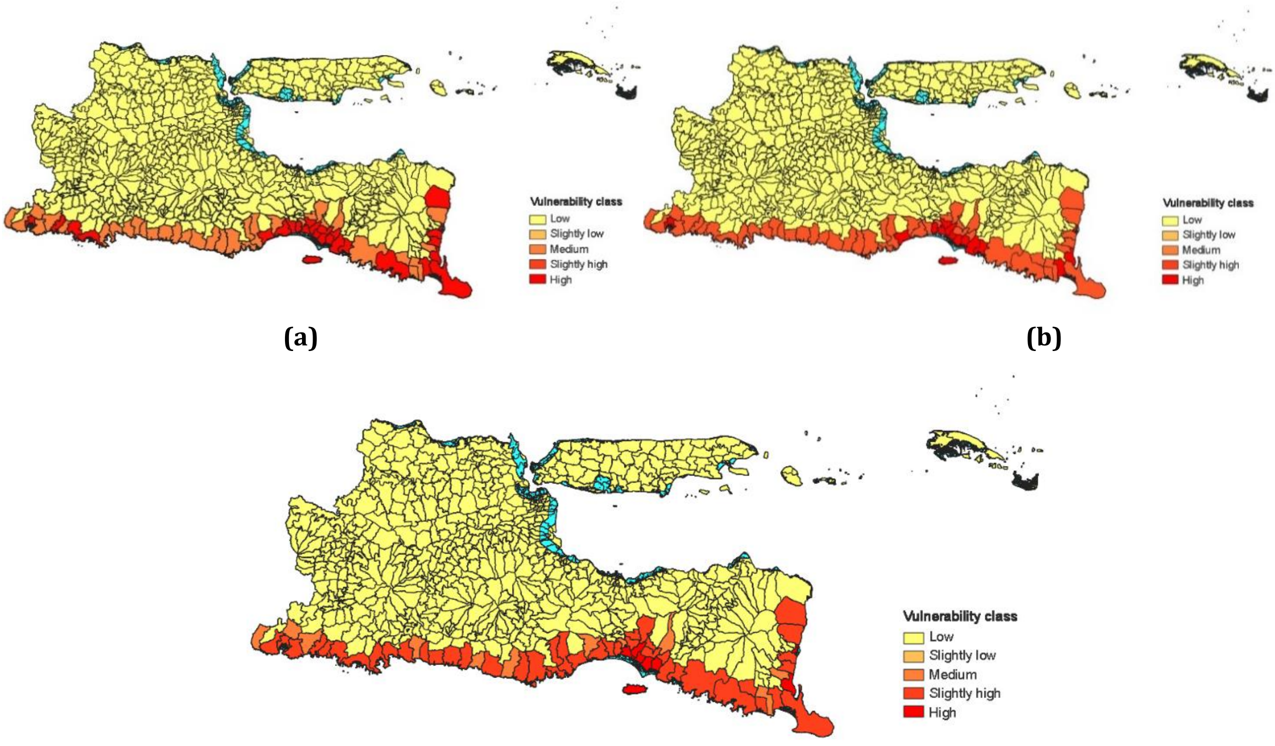


Fig. 5 *Comparative performance analysis*

The final stage of method evaluation, we conducted a performance comparison test between baseline methods, comparing Weighted Overlay classification and standard SVM with our proposed hybrid fuzzy-SVM method using an RBF kernel. The Weighted Overlay baseline used equal weighting (1/3 per index), based on the number of parameters used in this research: land cover index, elevation index, and inundation index. The SVM baseline used an RBF kernel with default C=1 and gamma=1/n_features, with additional tuning via grid search cross-validation.

The comparative analysis in Figure 5 reveals that the proposed Hybrid Fuzzy–SVM model achieved the highest overall accuracy (91.3%) and balanced performance across all metrics, with a precision of 0.911, recall of 0.910, and F1-score of 0.910. In contrast, the standard SVM model produced slightly lower results (accuracy 90.1%), while the traditional weighted overlay method achieved the lowest accuracy (87.6%). The chart visualization further confirms that the hybrid model consistently dominates across all evaluation dimensions, indicating its superior ability to capture non-linear relationships among tsunami-related variables, such as elevation, land cover, and inundation depth.



(a) **(b)**

(c)

Fig. 6 *Tsunami vulnerability mapping area based on (a) Weighted overlay; (b) Standard SVM; and (c) Hybrid fuzzy-SVM*

The other testing scenarios, we also visualized the tsunami vulnerability maps in this stage to demonstrate the performance differences among the three methods. The mapping visualization shows that both the weighted overlay method (a) and the standard SVM (b) exhibit poor transitional capability in classifying the five vulnerability classes, where the maps appear to display only contrasting two-color patterns that should have been represented with smoother gradations. In contrast, the Hybrid Fuzzy–SVM method with the best RBF model (c) demonstrates more adaptive and robust classification and mapping of tsunami vulnerability zones, as illustrated in Figure 6, and also provides a visualization of inundation under a predetermined scenario.

5. Conclusion

This study successfully demonstrated the integration of multi-source geospatial datasets and a hybrid fuzzy–Support Vector Machine (SVM) approach for fine-scale tsunami vulnerability mapping in the coastal region of East Java, Indonesia. The integration of fuzzy logic with SVM successfully addressed the limitations of conventional GIS-based weighted overlay and standard SVM models, particularly in handling the uncertainty and gradual transition between vulnerability levels. By introducing a fuzzy membership system to preprocess elevation, land cover, and inundation indices, the proposed model enabled each spatial unit to express partial membership across multiple vulnerability classes, thereby producing a smoother, more realistic vulnerability surface.

The proposed Hybrid model approach, employing both *Radial Basis Function (RBF)* and *Linear* kernels, proved effective in classifying tsunami vulnerability into five categories: *low*, *slightly low*, *medium*, slightly *high*, and *high*. All reported precision, recall, and F1-score values correspond to macro-averaged metrics computed under a one-vs-rest multi-class scheme. Among the two kernels, the RBF kernel achieved superior performance, with a macro-average accuracy of 91.3%, precision of 0.91, recall of 0.91, and F1-score of 0.91. The confusion matrix revealed minimal misclassification between adjacent classes, primarily within transition coastal areas with mixed elevation and heterogeneous land use. The ROC analysis under a One-vs-Rest configuration further validated the model's robustness, yielding AUC values between 0.88 and 0.96, with a macro-average AUC of 0.93. The hybrid fuzzy–SVM model also achieved superior classification performance compared to baseline methods, with higher overall accuracy, precision, recall, and F1-score values. These findings provide valuable insights for local disaster management agencies to strengthen early-warning systems and prioritize mitigation planning. Overall, this research demonstrates that integrating fuzzy logic into machine learning classifiers enhances model robustness, interpretability, and adaptability in complex coastal hazard mapping. The proposed hybrid fuzzy–SVM framework thus represents a significant methodological improvement in tsunami risk assessment, offering a replicable and data-driven model for other coastal regions worldwide.

Future research can explore several extensions of the proposed hybrid fuzzy–SVM framework. First, integrating additional geospatial parameters such as shoreline morphology, coastal slope, and population density could improve the model's predictive capability. Second, ensemble learning techniques (e.g., Random Forests, Gradient Boosting, or hybrid CNN–SVM architectures) may be tested to enhance accuracy and stability across spatial scales. Third, the model can be integrated into a real-time tsunami early-warning system using IoT-based sensors and near-real-time satellite data, such as Sentinel-1 SAR and MODIS thermal products. Finally, future work should also focus on developing adaptive fuzzy membership functions that dynamically adjust based on local geomorphological conditions. Such improvements will enable a more responsive and automated vulnerability assessment system that supports rapid decision-making for coastal resilience and disaster risk reduction.

Acknowledgement

The author would like to thank the Directorate of Research and Community Service (DPPM), Ministry of Higher Education, Science, and Technology, for funding support through the Fundamental Research Grant, under Master Contract Number 064/C3/DT.05.00/PL/2025.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Adi Susilo, Muhammad Fathur Rouf Hasan;* ***data collection:*** *Renaldi Primaswara Prasetya, Rony Prianto Nugraha;* ***analysis and interpretation of results:*** *Rizal Setya Perdana, Azizul Azhar Ramli;* ***draft manuscript preparation:*** *Muhammad Fathur Rouf Hasan, Renaldi Primaswara Prasetya. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Benchekroun, S., Omira, R., Baptista, M. A., El Mouraouah, A., Brahim, A. I., & Toto, E. A. (2015). Tsunami impact and vulnerability in the harbor area of Tangier, Morocco. *Geomatics, Natural Hazards and Risk*, 6(8), 718–740. https://doi.org/10.1080/19475705.2013.858373.

[2] Benazir, & Oktari, R. S. (2024). Assessing tsunami risk along the Aceh coast, Indonesia: a quantitative analysis of fault rupture potential and early warning system efficacy for predicting arrival time and flood extent. *Natural Hazard*, 120, 4875–4900. https://doi.org/10.1007/s11069-02406401-x.

[3] Isdianto, A., Kurniasari, D., Subagiyo, A., Haykal, M. F., & Supriyadi, S. (2021). Pemetaan Kerentanan Tsunami untuk Mendukung Ketahanan Wilayah Pesisir. *Jurnal Permukiman*, *16*(2), 90–100. https://doi.org/10.31815/jp.2021.16.90-100

[4] Hou, J., Gao, Y., Fan, T., Wang, P., Wang, Y., Wang, J., & Lu, W. (2023). Tsunami Risk Change Analysis for Qidong County of China Based on Land Use Classification. *Journal of Marine Science and Engineering*, *11*(2), 379. https://doi.org/10.3390/jmse11020379

[5] Wibowo, G. C. A., Prasetyo, S. Y. J., & Sembiring, I. (2023). Tsunami Vulnerability and Risk Assessment in Banyuwangi District using machine learning and Landsat 8 image data. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *22*(2), 365-380. https://doi.org/10.30812/matrik.v22i2.2677

[6] Liu, C. M., Rim, D., Baraldi, R., & LeVeque, R. J. (2021). Comparison of machine learning approaches for tsunami forecasting from sparse observations. *Pure and Applied Geophysics*, *178*(12), 5129-5153. https://doi.org/10.1007/s00024-021-02841-9

[7] Wieland, M., Liu, W., & Yamazaki, F. (2016). Learning Change from Synthetic Aperture Radar Images: Performance Evaluation of a Support Vector Machine to Detect Earthquake and Tsunami-Induced Changes. *Remote Sensing*, *8*(10), 792. https://doi.org/10.3390/rs8100792

[8] Lessy, M. R., & Sabar, M. (2021). Mapping Tsunami Vulnerability Area for Bacan Sub-District and Its Surroundings – North Maluku Province. *E3S Web Conf.*, *328*, 04024. https://doi.org/10.1051/e3sconf/202132804024.

[9] Ghadamode, V., Srivastava, K., Singh, R. K., & Pandey, A. K. (2022). Spatial analysis techniques for tsunami vulnerability and inundation mapping of Andaman region using remote sensing, GIS, AHP, and Fuzzy logic methods. *Environmental Earth Sciences*, *81*(17), 427. https://doi.org/10.1007/s12665-022-10548-w

[10] Sukmana, H., Durachman, Y., Amri, A., & Supardi, S. (2024). Comparative Analysis of SVM and RF Algorithms for Tsunami Prediction: A Performance Evaluation Study. *Journal of Applied Data Sciences, 5*(1), 84-99. https://doi.org/10.47738/jads.v5i1.159

[11] Strunz, G., Post, J., Zosseder, K., Wegscheider, S., Mück, M., Riedlinger, T., Mehl, H., Dech, S., Birkmann, J., Gebert, N., Harjono, H., Anwar, H. Z., Sumaryono, Khomarudin, R. M., & Muhari, A. (2011). Tsunami risk assessment in Indonesia. *Natural Hazards and Earth System Sciences*, *11*(1), 67–82. http://doi.org/10.5194/nhess-11-67-2011

[12] Prasetyo, H., & Maulana, A. E. (2025). Penerapan Algoritma Support Vector Machine untuk Melakukan Prediksi Potensi Tsunami di Indonesia. *MULTINETICS*, *10*(2), 174–182. https://doi.org/10.32722/multinetics.v10i2.7245

[13] Suppasri, A., Fukutani, Y., Abe, Y., & Imamura, F. (2013). Relationship between earthquake magnitude and tsunami height along the Tohoku coast based on historical tsunami trace database and the 2011 Great East Japan Tsunami. *Report of Tsunami Engineering*, *30*, 37-49.

[14] Isnaeni, A. Y., & Prasetyo, S. Y. J. (2022). Klasifikasi Wilayah Potensi Risiko Kerusakan Lahan Akibat Bencana Tsunami Menggunakan Machine Learning. *Jurnal Teknik Informatika Dan Sistem Informasi*, *8*(1), 33. https://doi.org/10.28932/jutisi.v8i1.4056

[15] Kimpton, T., Higuera, P., Whittaker, C., Wotherspoon, L., & Zorn, C. (2024). A rapid simplified method for determining tsunami inundation extent based on energy conservation. *Computer-Aided Civil and Infrastructure Engineering*, 39, 1811–1829. https://doi.org/10.1111/mice.13168

[16] Tika, N., Rakyan Paksi, N., & Muhammad, D. (2025). Advancing Tsunami Vulnerability Modelling: A Systematic Review and Bibliometric Analysis of Remote Sensing and GIS Applications. *European Journal of Geography*, *16*(2), 298–318. https://doi.org/10.48088/ejg.n.tik.16.2.075.095

[17] Mardiani, E., Rahmansyah, N., Ningsih, S., Lantana, D. A., Wulandana, N. P., Lombu, A. A., & Budyarti, S. (2024). Classification of Potential Tsunami Disaster Due to Earthquakes in Indonesia Based on Machine Learning. *International Journal Software Engineering and Computer Science (IJSECS)*, *4*(1), 13–23. http://doi.org/10.35870/ijsecs.v4i1.2084

[18] Siregar, A., Sudianto, F., & Nooraeni, R. (2024). Prediction of Tsunamis in Indonesia Using an Optimized Neural Network with SMOTE. *SISTEMASI, 13*(5), 1914-1925. https://doi.org/10.32520/stmsi.v13i5.4263

[19] Sambah, A. B., Miura F., Guntur, & Fuad. (2018). Integrated Satellite Remote Sensing and Geospatial Analysis for Tsunami Risk Assessment. *GEOMATE Journal*, *14*(44), 96–101. https://doi.org/10.21660/2018.44.7127

[20] Guntur, Sambah, A. B., Miura, F., Fuad, & Arisandi, D. M. (2016). Assessing Tsunami Vulnerability Areas using Satellite Imagery and Weighted Cell-Based Analysis. *GEOMATE Journal*, *12*(34), 115–122. http://dx.doi.org/10.21660/2017.34.2726

[21] Sambah, A. B., Masnagari, L. M. S., Fuad, M A. Z., & Intyas, C. A. (2024). Tsunami Run-Up Modelling in Comparison with Coastal Vulnerability Mapping. *GEOMATE Journal*, *27*(119), 10–17. https://doi.org/10.21660/2024.119.4000

[22] Rehman, A., & Zhang, H. (2025). Tsunami Vulnerability Analysis of Makran Subduction Zone through Fuzzy Logic. *Indonesian Journal on Geoscience*, 12(3), 413-421. https://doi.org/10.17014/ijog.12.3.413-421

[23] Jumadi, J., Priyono, K. D., Amin, C., Saputra, A., Gomez, C., Lam, K. C., Rohman, A., Patel, N., Sattar, F., Nawaz, M. & Wardani, K. S. (2025). Tsunami Risk Mapping and Sustainable Mitigation Strategies for Megathrust Earthquake Scenario in Pacitan Coastal Areas, Indonesia. *Sustainability*, *17*(6), 2564. https://doi.org/10.3390/su17062564

[24] Supendi, P., Sianipar, D., Widiyantoro, S., Rawlinson, N., Daryono, Prayitno, B. S., & Rahayu, T. (2022). Analysis of the April 10, 2021 (Mw 6.1) destructive intra-slab earthquake, East Java, Indonesia. *Physics of the Earth and Planetary Interiors*, *326*, 106866. https://doi.org/10.1016/j.pepi.2022.106866

[25] Rochman, A, D., Irawan, D. E., Sumintadireja, P., & Ramadhan, G. (2024), *Data For: Bandung Basin Digital Elevation Model.* https://data.mendeley.com/datasets/h4tfpp7xyr/1 accessed on September 30, 2025

[26] Berryman, K. (2005). *Review of Tsunami Hazard and Risk in New Zealand, Report by the Institute of Geological and Nuclear Sciences*, Lower Hutt, New Zealand: Institute of Geological & Nuclear Sciences.

[27] Pertiwi, I., Fattah, M., & Rauf, A. (2018). Estimation of Tsunami Inundation and Disaster Mitigation in Bulukumba, Indonesia. *Jurnal Geofisika, 16*(1), 1-8. https://doi.org/10.36435/jgf.v16i1.13

[28] Prasetya, G., Borrero, J., de Lange, W., Black, K., & Healy, T. (2011). Modeling of inundation dynamics on Banda Aceh, Indonesia during the great Sumatra tsunamis December 26, 2004. *Natural hazards*, *58*(3), 1029-1055. https://doi.org/10.1007/s11069-010-9710-7

[29] Widiyanto, W., Santoso, P. B., Hsiao, S. C., & Imananta, R. T. (2019). Post-event field survey of September 28 2018 Sulawesi earthquake and tsunami. *Natural Hazards and Earth System Sciences*, *19*(12), 2781-2794. https://doi.org/10.5194/nhess-19-2781-2019, 2019.

[30] Yousif, Y. K., Bermani, A. K., Aldulaimi, M. H., Khalaf, M., Mohammed, R. B., & Almihi, A. J. (2025). A Fuzzy-based Cluster Head Selection Technique for Optimizing Communication of VANETs. *Journal of Soft Computing and Data Mining*, *6*(1), 127-137. https://doi.org/10.30880/jscdm.2025.06.01.009

[31] Kusnaidi, M. R., Gulo, T., & Aripin, S. (2022). Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode KMeans Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal. *Journal of Computer System and Informatics (JoSYC), 3*(4), 330–338. https://doi.org/10.47065/josyc.v3i4.211 2

[32] Ali, A. M., & Mohammed, M. A. (2024). Optimized Cancer Subtype Classification and Clustering Using Cat Swarm Optimization and Support Vector Machine Approach for Multi-Omics Data. *Journal of Soft Computing and Data Mining*, *5*(2), 223-244. https://doi.org/10.30880/jscdm.2024.05.02.017

[33] Mostafa, S. A., Hasan, D. F., Abdul-Jabar, M. A., & Mustapha, A. (2025). An Ensemble Learning Model for Multi-Type Cancer Prediction in Clinical Diagnostic Decision Support Systems. *Journal of Soft Computing and Data Mining*, *6*(1), 230-246. https://doi.org/10.30880/jscdm.2025.06.01.015

[34] Redyansyah, R., Satriadi, A., & Saputro, S. (2017). Pemodelan Penjalaran Gelombang Tsunami dan Analisa Daerah Jangkauan di Teluk Sumbreng, Trenggalek. *Journal of Oceanography, 6*(1), 295 – 304.

[35] Sambah, A. B., Miura, F., Guntur, Sunardi, & Febriana A. F. (2019). Geospatial Model of Physical and Social Vulnerability for Tsunami Risk Analysis. *GEOMATE Journal*, *17*(63), 29–34. https://doi.org/10.21660/2019.63.4684

[36] Widada, S., Darda, I. M., & Satriadi, A. (2022). Identifikasi Wilayah Terdampak Tsunami Berdasarkan Peta Ancaman Tsunami di Kabupaten Lumajang, Jawa Timur. *Buletin Oseanografi Marina, 11*(3), 291-305. https://doi.org/10.14710/buloma.v11i3.44032