PAN
POLSKA AKADEMIA NAUK

JOURNAL OF WATER AND LAND DEVELOPMENT

e-ISSN 2083-4535

ITP-PIB

Polish Academy of Sciences (PAN) Institute of Technology and Life Sciences – National Research Institute (ITP – PIB)

JOURNAL OF WATER AND LAND DEVELOPMENT
DOI: 10.24425/jwld.2026.158720
2026, No. 69 (IV–VI): 225–234

Random forest-based prediction of flash flood susceptibility under land-use dynamics in East Java watersheds

Entin Hidayah*[1)], Jojok Widodo Soetjipto[1)], Juliastuti Juliastuti[2)], Azmeri Azmeri[3)], Alexander Agung Santoso Gunawan[4)], Joko Widodo[5)], Josaphat Tetuko Sri Sumantyo[6)], Afif Amiluddin[7)]

[1)] Universitas Jember, Faculty of Engineering, Civil Engineering Department, Jl. Kalimantan No. 37, Tegalboto Sumbersari, Kec. Sumbersari, Kabupaten Jember, Jawa Timur 68121, Indonesia

[2)] Bina Nusantara University, Faculty of Engineering, Civil Engineering Department, Jl. K. H. Syahdan No. 9, Kemanggisan, 11480 Palmerah Jakarta, Indonesia

[3)] Universitas Syiah Kuala, Faculty of Engineering, Syech Abdur-Rauf No. 7 Darussalam, 23111, Banda Aceh, Indonesia

[4)] Bina Nusantara University, Computer Science Department, School of Computer Science, Jl Kebon Jeruk Raya no 27, 11530 Jakarta Barat, Indonesia

[5)] National Research and Innovation Agency (BRIN), Research Center for Geoinformatics, Gedung B.J. Habibie, Jl. M.H. Thamrin No. 8, Jakarta, 10340, Indonesia

[6)] Chiba University, Center for Environmental Remote Sensing, 1-33 Yayoi-cho, Inage-ku, 263-8522 Chiba-shi, Japan

[7)] Universitas Mochammad Sroedji, Faculty of Engineering, Civil Engineering Department, Jalan Sriwijaya No. 32, Sumbersari District, 68121 Jember, Indonesia

* Corresponding author

RECEIVED 20.09.2025 ACCEPTED 20.02.2026 AVAILABLE ONLINE 16.06.2026

Abstract: The frequent occurrence of flash floods in the small mountainous areas around Argopuro, Ijen and Raung in East Java, with steep terrain and short hydrological response times, poses significant challenges for risk assessment in regions with limited hydrological observations. This study assesses flash-flood susceptibility in eight small watersheds in East Java, Indonesia, using a random forest (RF)-based approach. A flash-flood inventory for 2012–2022 was compiled and used to train and validate the model. Ten conditioning factors were derived from a Digital Elevation Model (DEM), river-network data and satellite imagery, including slope, topographic wetness index (*TWI*), topographic position index (*TPI*), river density, land use and normalised difference vegetation index (*NDVI*). Land-use changes from 2001 to 2020 were used to assess their influence. To strengthen the hydrological context, event-based rainfall data corresponding to flash-flood occurrences and field-based photographic documentation were incorporated. The accuracy of the RF model was assessed using ROC curves and root mean square error (*RMSE*), indicating outstanding performance (scores of 0.96–0.99). Based on the model results, four main factors were identified as key determinants of flash-flood occurrence: land cover and its changes, slope steepness, river-network density and *TPI*. The spatial patterns generated by the RF model delineate areas with varying susceptibility levels, providing actionable information for prioritising mitigation measures. These findings contribute to a better understanding of flash-flood drivers in small watersheds and provide a scientific basis for local governments and stakeholders to design targeted risk-reduction strategies that combine land-use planning, watershed management and community-based adaptation.

Keywords: flash flood, geomorphology, land-use change, random forest, susceptibility

INTRODUCTION

Flash floods have increasingly affected East Java over the last decade, with more than 25 documented events between 2012 and 2024 across five districts: Banyuwangi, Jember, Bondowoso, Situbondo and Probolinggo. These events, concentrated in small mountainous watersheds, have caused fatalities, damaged hundreds of hectares of rice fields and inundated settlements, schools, fishponds and critical infrastructure. Unlike ordinary floods, flash floods occur suddenly after short but intense rainfall and often transport large amounts of debris, which amplifies their destructive impacts. They have ruptured levees, damaged irrigation systems, destroyed bridges and disrupted regional connectivity. Their localised nature and rapid onset make them highly unpredictable, posing serious challenges for disaster preparedness and mitigation (Hocini *et al*., 2021; Yuan *et al*., 2021).

Small watersheds in upland areas are especially sensitive to flash floods because of their hydrological and topographic settings. Intense precipitation generates higher runoff volumes in mountainous regions than in lowlands (Sharifi, Eitzinger and Dorigo, 2019). Steep slopes accelerate water flow, while rugged terrain channels runoff rapidly downstream, threatening riverbank stability (Choi *et al*., 2020). Contributing factors include inadequate drainage systems, altered flow routes, poorly managed levees and extensive land-use changes that reduce infiltration capacity (Du, Fan and Pu, 2020; Zhang *et al*., 2021). Climate change further intensifies these risks by increasing rainfall variability and the frequency of extreme weather events (Abid *et al*., 2021). Given the complexity of these interacting drivers, identifying and understanding the environmental factors that condition flash-flood occurrence is critical for developing effective prevention and mitigation strategies.

Recent studies have adopted various techniques to assess flash-flood susceptibility. Traditional approaches include multi-criteria decision-making (Tariq *et al*., 2022) and bivariate statistical models (Hidayah, Halik and Widiarti, 2023). Advances in computational methods have introduced machine-learning and deep-learning algorithms, which offer greater accuracy and flexibility in handling non-linear relationships (Chen *et al*., 2023). Among these, the random forest (RF) algorithm has emerged as one of the most effective tools for susceptibility mapping. RF models have achieved prediction accuracies exceeding 80% in flash-flood modelling and even higher rates in flood-forecasting applications (Mirzaei *et al*., 2021). Their capacity to rank the importance of conditioning factors provides additional insight beyond susceptibility mapping, establishing RF as a state-of-the-art method in hazard assessment (Abedi *et al*., 2022).

Diverse environmental and morphometric variables shape flash-flood susceptibility, and their relative influence varies across regions. Studies have highlighted the importance of land use/land cover (LULC), vegetation indices such as *NDVI*, precipitation and proximity to rivers (Cao *et al*., 2020; Choudhury *et al*., 2022). Other parameters, such as slope, aspect, elevation, river density and indices derived from digital elevation models (e.g. topographic wetness index – *TWI*, convergence index – *CI* and topographic position index – *TPI*), also significantly influence runoff generation and infiltration (Sellami, Maanan and Rhinane, 2022). In particular, land-use patterns, surface slope, terrain position (as represented by *TPI*) and river density have frequently emerged as key contributors to flash-flood susceptibility, as they directly control flow concentration, runoff pathways and landscape connectivity (Sellami, Maanan and Rhinane, 2022). For instance, deforestation and the conversion of natural vegetation into agricultural or built-up land increase runoff and flood volume, whereas afforestation can reduce peak flows (Bathurst *et al*., 2020). The *NDVI* is an effective indicator of vegetation health and greenness, with lower values indicating reduced canopy cover and increased flood susceptibility (Sidi *et al*., 2021). Despite extensive research on flash-flood susceptibility, key limitations persist. Many studies focus on single watersheds, limiting the generalisation of conditioning-factor importance across diverse land-water systems and geomorphological settings. Moreover, land use/land cover and vegetation indicators are often treated as static variables, neglecting the long-term land-cover dynamics that influence runoff.

To address these limitations, this study adopts a multi-watershed RF-based framework that simultaneously analyses eight small watersheds in East Java with documented flood events. By integrating multi-temporal land use/land cover datasets from 2001 to 2020, the proposed approach captures long-term land-cover dynamics and enables the identification of transferable conditioning-factor patterns beyond single-watershed and static-variable analyses. The objectives are threefold: (1) to evaluate the role of topographic, hydrological and land-use variables in determining flash-flood susceptibility; (2) to analyse the influence of land-cover change on susceptibility patterns; and (3) to generate susceptibility maps that can guide mitigation strategies. The findings advance the state of the art in flash-flood modelling and offer practical implications for watershed management and disaster risk reduction in vulnerable mountainous regions of Indonesia.

MATERIALS AND METHODS

STUDY AREA

The study area is characterised by complex mountainous terrain influenced by three major volcanic massifs: Mount Argopuro, Mount Ijen and Mount Raung. For analytical purposes, the eight mountainous catchments were classified according to elevation zones, with five (Wonoboyo, Curah Nongko, Pecaron, Deluwang and Gelondong) located below 500 m above sea level (m a.s.l.) and three (Banyuputih, Pekalen and Klokoran) situated between 500 and 1000 m a.s.l. (Fig. 1). A dataset of documented flash-flood events during the period 2008–2024 was compiled and used to characterise flood-prone catchments, with individual mountainous catchments experiencing between one and six recorded events (Tab. S1–S7).

These flash-flood events were consistently associated with intense rainfall recorded at multiple rain-gauge stations surrounding the affected sub-watersheds, with daily rainfall totals commonly exceeding 50 mm per day, corresponding to the BMKG classification of heavy to very heavy rainfall (BMKG, 2023). The consistency of rainfall magnitudes across neighbouring stations indicates that the events were driven by widespread precipitation rather than isolated local storms.

Field documentation further supports the severity of these events. Photographs taken during and immediately after flood occurrences show channel overtopping, sediment-laden flows, bank erosion and infrastructure damage (Tab. S1–S7). Together,

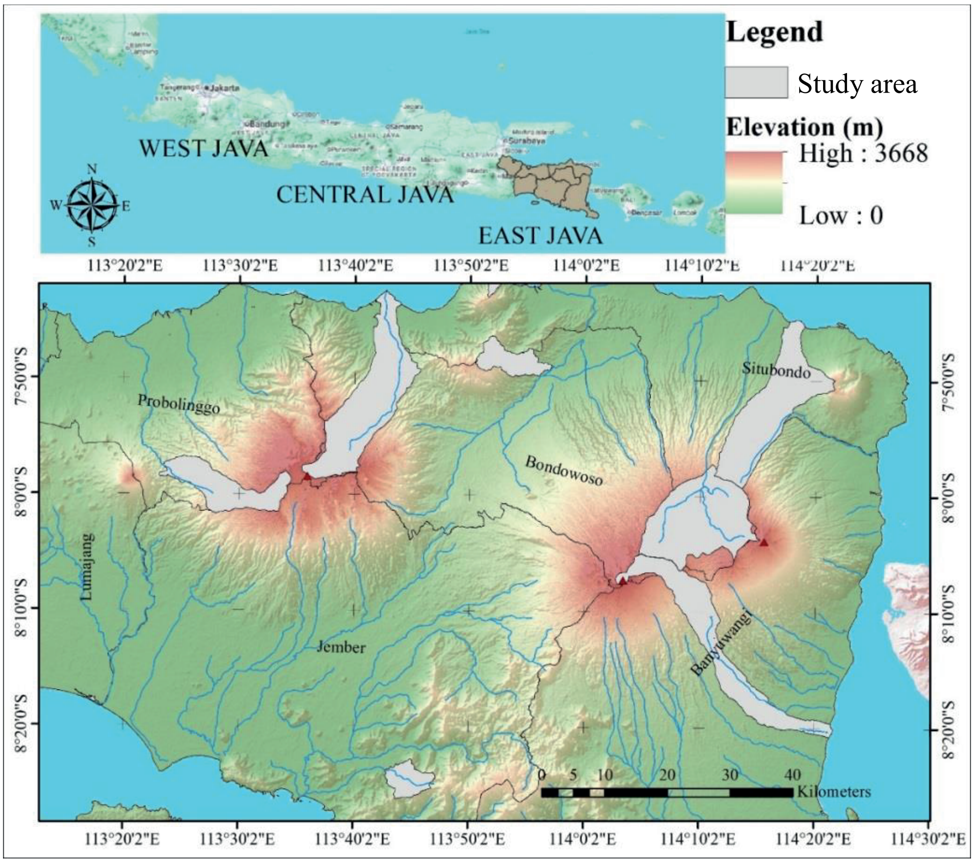

Fig. 1. Research area; source: own elaboration

the multi-station rainfall measurements and visual field evidence provide independent confirmation that the documented flash floods were triggered by intense rainfall over steep mountainous catchments.

DATA SOURCES AND SPATIAL PROCESSING

The methodology, summarised schematically in Figure 2, involved three main steps: (1) selecting and classifying factors influencing flash-flood occurrence at eight locations; (2) estimating susceptibility-index values using an RF model and validation; and (3) analysing spatial patterns among influential factors.

Data were obtained from four primary sources: field surveys, a Digital Elevation Model (DEM), the river network and satellite imagery. Flood-inventory data were used for training and testing. The 30 × 30 m DEM (USGS SRTM) was processed to derive slope, aspect, topographic position index (*TPI*), topographic wetness index (*TWI*), convergence index (*CI*) and curvature parameters. These parameters represent terrain controls on runoff generation, flow concentration and saturation potential.

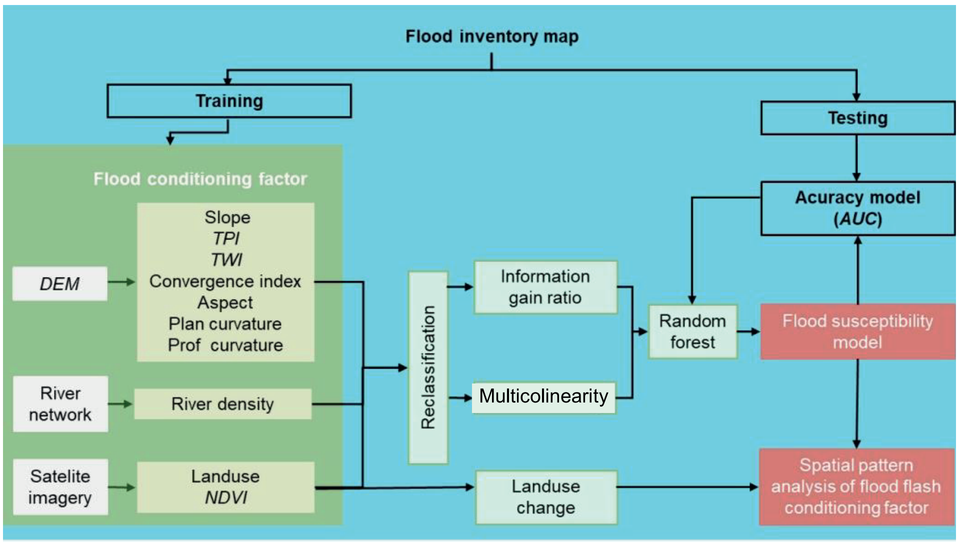

Fig. 2. Workflow of methodology; *TPI* = topographic position index, *TWI* = topographic wetness index, *NDVI* = normalised difference vegetation index; source: own elaboration

River density was extracted from a 1:25,000 topographic map. Land-cover information was derived from Landsat imagery (2001 and 2020) using supervised classification and was validated with field surveys. Eight land-cover classes were identified: forest, plantation, settlement, paddy fields, shrubs, coastal sand, river and pond. The *NDVI* derived from Sentinel-2 (2020) was also used, focusing on dry-season imagery to capture minimum vegetation cover before rainfall. All spatial data were processed in QGIS and standardised into raster grids (30 m resolution) to support random forest (RF)-based susceptibility mapping.

FLASH FLOOD CONDITIONING FACTORS

The selection of conditioning factors is essential for assessing flash-flood susceptibility. Morphometric parameters such as slope, aspect, *TPI*, *TWI* and *CI* strongly influence flood potential. Steeper slopes accelerate runoff, aspect affects local hydrological conditions, and *TPI* reflects terrain position relative to neighbouring areas, while high *TWI* and convergence values indicate saturated zones with greater runoff and limited infiltration (Saleh, Yuzir and Sabtu, 2022). River density also plays a key role, as areas near dense river networks are more prone to floods (Tariq *et al.*, 2022). Land use and *NDVI* significantly influence the occurrence of flash floods. Less permeable land-use types, such as densely built-up areas, generate higher runoff than forests or cropland because the surface absorbs less water and rainfall is conveyed more quickly into drainage channels, increasing flash-flood risk. Each land-use class affects flood events differently (Chen *et al.*, 2020). The conversion of natural vegetation into farmland, the expansion of built-up areas and the loss of forest can increase runoff and flood volume, while afforestation may reduce the frequency of peak discharge (Costache *et al.*, 2020). The *NDVI* indicates vegetation greenness, with decreasing values suggesting a higher potential for flash floods. This index is used to detect land-cover change because it shows positive values for vegetation, values close to zero for bare soil and negative values for water (Sidi *et al.*, 2021). Areas with lower *NDVI* values are therefore more sensitive to flash floods (Saha, Gayen and Bayen, 2022). To support flash-flood inventory validation, event-based rainfall data were compiled for each small watershed (Tabs. S1–S7).

MULTICOLLINEARITY AND INFORMATION GAIN RATIO

The selection of factors is a crucial phase in preprocessing, as it aims to improve data quality by reducing redundancy while preserving variables with strong discriminative capability (Ahmad and Aziz, 2019). Multicollinearity, defined as a strong linear relationship among explanatory variables, can obscure the relative importance of factors and reduce the interpretability of spatial susceptibility models.

To address this issue, a correlation-based multicollinearity analysis was performed by calculating pairwise Pearson correlation coefficients among all conditioning factors. A threshold of $|r| \geq 0.80$ was adopted to identify strongly correlated variables, following common practice in GIS-based flood susceptibility studies. When this threshold was exceeded, only one variable was retained based on its physical relevance and contribution to flash flood processes.

Furthermore, the importance of the selected conditioning factors was evaluated using the gain ratio (*GR*) method, a supervised feature-selection technique derived from information theory that normalises information gain to reduce bias towards attributes with many distinct values (Arora and Kaur, 2020). The *GR* is defined in Equation (1).

$$GR\ (\mathrm{X}) = \frac{\mathrm{Gain}(X)}{\mathrm{SplitInfo}(X)} \tag{1}$$

where: SplitInfo(X) = the entropy of the probability distribution after data partitioning, as shown in Equation (2).

$$\mathrm{SplitInfo}(X) = -\Sigma\left(\frac{|\mathrm{Mi}|}{\mathrm{M}}\right)\log_2\left(\frac{|\mathrm{Mi}|}{|\mathrm{M}|}\right) \tag{2}$$

where: $|Mi|$ = the number of members of the Mi subset of the M training data.

RANDOM FOREST

Random forest machine learning was used as the classification method for flash-flood events. The random forest (RF) is a classification method that consists of a combination of decision trees (Abedi *et al.*, 2022). The method combines bootstrap aggregation and randomisation in node selection during the construction of decision trees. A decision tree with L leaves divides the feature space into L parts, Rl, $1 \leq l \leq L$. For each decision tree, the prediction function follows Equations (3) and (4) (Chen *et al.*, 2023).

$$f(x) = \sum_{l=1}^{L} c_l \Pi(x, Rl) \tag{3}$$

$$\Pi(x, Rl) = \begin{cases} 1, & if\ x \in Rl \\ 0, & otherwise \end{cases} \tag{4}$$

where: L = the number of parts of the feature space, Rl = the l-th part and c_l is the constant corresponding to the l-th part.

The final classification is based on the majority vote across all trees.

Random forest parameters were adopted from commonly used settings reported in previous flood-susceptibility studies to ensure model stability and consistency. No extensive hyperparameter sensitivity analysis was performed, as the primary objective of this study is spatial susceptibility mapping rather than algorithm optimisation.

Model performance was evaluated using a confusion matrix derived from the training and testing results, summarising outcomes as true positives (*TP*), false positives (*FP*), true negatives (*TN*) and false negatives (*FN*) (Kamarudin, Cox and Kolamunnage-Dona, 2017). Although misclassifications are unavoidable, effective models exhibit low and balanced error rates. Performance was assessed using accuracy, recall (true positive rate, *TPR*) and precision, with formulations given in Equations (5)–(7).

$$\mathrm{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

Furthermore, the receiver operating characteristic (*ROC*) curve was employed to evaluate prediction quality by plotting the false positive rate (*FPR*) against the true positive rate (*TPR*) (Wang *et al*., 2022). The area under the curve (*AUC*) provides a quantitative measure of overall model performance, with higher values indicating stronger discriminative ability (Yao *et al*., 2022). To enhance validation, *ROC* curves were generated for different subgroups of flash-flood events. Since *ROC* analysis alone does not define explicit susceptibility thresholds, the resulting susceptibility map was classified into five equal-area susceptibility classes. This classification facilitates spatial interpretation while preserving model accuracy in distinguishing flood-prone and non-flood areas.

RESULTS AND DISCUSSION

FLASH FLOOD CONDITIONING FACTOR RESULTS

The resulting correlation matrix is provided in Table S8 to enable independent evaluation of inter-variable relationships. The analysis indicates that no significant multicollinearity is present; therefore, all ten variables were retained and used in the modelling procedure: *NDVI*, plan curvature, aspect, profile curvature, convergence, *TWI*, slope, land use, *TPI* and river density. This result confirms that each factor contributes unique information to the random forest (RF) classifier and avoids redundancy that could obscure interpretation.

The relative importance of these variables was evaluated using the gain ratio (*GR*) based on average merit values, providing an interpretable measure of their contribution to flash-flood susceptibility across the eight watersheds. Across the watersheds, land use consistently exhibits the highest average merit, followed by slope, river density and *TPI* (Fig. 3). The dominance of land use, slope, river density and *TPI* reflects their complementary roles in controlling watershed hydrological response. Land use governs runoff generation by regulating infiltration and surface storage, commonly represented through variations in curve number (Hu, Fan and Zhang, 2020). Slope controls flow velocity and response time by shortening the time of concentration and amplifying peak discharge (Sultan *et al*., 2022). River density determines the efficiency of runoff conveyance by accelerating flow concentration towards the channel network (Ogden *et al*., 2011), while *TPI* represents topographic convergence and divergence that regulate the spatial distribution and accumulation of runoff (Chowdhury, 2023). Together, these factors control both the magnitude and timing of watershed response to intense rainfall, explaining their dominant influence on flash-flood susceptibility.

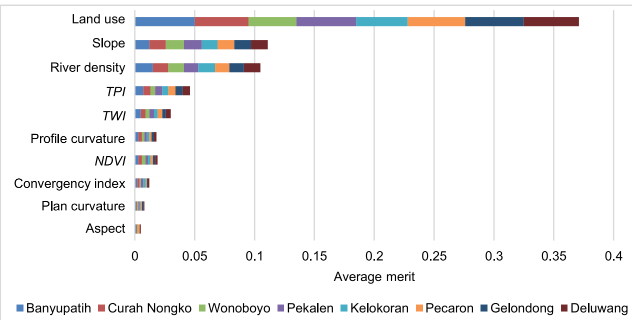

Fig. 3. Overall average merit values of the gain ratio; *TPI* = topographic position index, *TWI* = topographic wetness index, *NDVI* = normalised difference vegetation index; source: own study

FLASH FLOOD SUSCEPTIBILITY PREDICTION USING RANDOM FOREST

Flash-flood susceptibility was classified into five levels, ranging from very low to very high. Across the eight watersheds, areas classified as very high susceptibility account for approximately 58–77%, while the high and medium susceptibility classes account for 31–54% and 11.6–24.0%, respectively (Fig. 4). Across the eight watersheds, flash-flood susceptibility decreases sequentially from Gelondong to Klokoran, with higher-ranked watersheds characterised by larger proportions of high to very high susceptibility classes and lower-ranked watersheds dominated by low to medium classes. This pattern indicates pronounced spatial variability in flash-flood susceptibility, highlighting clear differences in relative risk among the analysed catchments. Model performance was evaluated using receiver operating characteristic (*ROC*) and root mean square error (*RMSE*) metrics for both the training and testing phases. Training *ROC* values range from 0.87 to 0.99, while testing *ROC* values remain consistently high (≥0.83), indicating strong discriminative capability across all watersheds. Low *RMSE* values further confirm the robustness and generalisation ability of the random forest model. Similar evaluation frameworks using *ROC* and *RMSE* are widely adopted in flash-flood susceptibility studies to assess spatial classification accuracy and model reliability (Nguyen *et al*., 2020; Shirzadi *et al*., 2020). Detailed performance metrics for each watershed are provided in Table S9.

LAND COVER CHANGE VERSUS FLASH FLOODS

Land cover is classified into eight categories according to Yang *et al.* (2012). Percentage changes in land-use/land-cover (LULC) classes between 2001 and 2020 for each watershed are presented

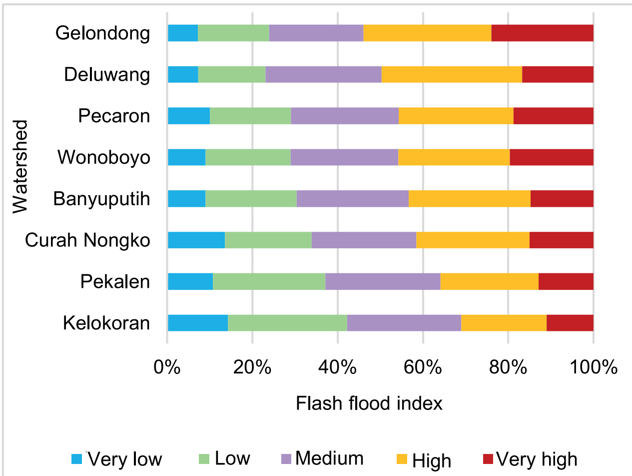

Fig. 4. Flash flood index; source: own study

in Figure S1, where negative values indicate land-cover loss and positive values indicate gains in areal proportion. Hydrologically, land cover influences runoff response through variations in *CN*, which integrates vegetation condition, soil properties and land management (Abdelkebir, Mokhtari and Engel, 2024).

Over the 20 years from 2001 to 2020, the Banyuputih, Curah Nongko, Pekalen, Pecaron and Deluwang watersheds experienced a decline in forest cover, which was replaced by plantations or paddy fields. Forest cover decreased markedly in Banyuputih (–38.0%) and Pecaron (–66.0%), while plantation areas increased significantly, reaching +34.0% in Banyuputih, +42.0% in Deluwang and up to +72.0% in Klokoran. Paddy fields also expanded notably in Deluwang (+24.8%). These transitions are hydrologically important because forests generally exhibit lower CN values. In contrast, plantations and paddy fields tend to increase CN due to soil compaction and reduced surface retention, promoting higher runoff and peak discharge (Abdelkebir, Mokhtari and Engel, 2024; Widyasasi *et al.*, 2024).

In contrast, the Wonoboyo, Klokoran and Gelondong watersheds did not show substantial land-use change. These areas are mainly covered by scrubland and contain only small patches of forest, indicating that flash floods are not caused solely by deforestation but can also be associated with the prevalence of shrubland cover. Flash-flood susceptibility can therefore occur in shrubland or agricultural watersheds where infiltration capacity and surface roughness are low. Shrubland- and agriculture-dominated watersheds can exhibit rapid runoff responses under intense rainfall because of inherently low infiltration capacity and surface roughness (Mansida *et al.*, 2025).

The *GR* analysis (Fig. 3) indicates that land use is the most informative conditioning factor for distinguishing flash-flood-prone areas across the studied watersheds. Rather than implying a direct causal relationship, this result suggests that land-use classes integrate multiple surface characteristics, including vegetation condition, topographic condition and long-term land-management practices. Changes in land-use composition, such as forest conversion, plantation expansion and the dominance of shrubland or agricultural land, provide strong discriminatory information for flash-flood susceptibility at the watershed scale.

Although forest cover can reduce runoff and attenuate flood peaks through increased interception, infiltration and evapotranspiration, its influence on flash floods induced by short-duration, high-intensity rainfall may be limited (Herath *et al.*, 2025). This supports observations that flash floods occur even in watersheds with limited forest change or shrubland dominance, where intense rainfall overwhelms hydrological buffering capacity. The LULC analysis focuses on quantified areal transitions relevant to susceptibility modelling and does not aim to perform detailed statistical testing, fragmentation metrics or change-rate intensity analysis.

GEOMORPHOLOGICAL CHARACTERISTICS AFFECTING FLASH FLOODS

Slope

The slope gradient in the eight watersheds is categorised into five classes: [0°; 5°), [5°; 10°), [10°; 25°), [25°; 45°), and >45°. Slopes with gradients over 40% are present in four specific watersheds: Wonoboyo, Curah Nongko, Pekalen and Deluwang, as shown in Figure S2. These watersheds exhibit steep conditions, with slope angles ranging from 25° to 45°. The Gelondong watershed features significantly steeper slopes, exceeding 45°.

The Banyuputih and Pecaron watersheds have moderate inclines, ranging from 10° to 25°. In the Klokoran watershed, there are two main slope classes: moderately steep and very steep. Flood occurrence is strongly influenced by slope gradient, as steeper slopes produce more rapid and forceful flows. For instance, in hilly regions, steeper slopes can lead to larger and faster peak flows (Kowalczyk *et al.*, 2023). According to previous studies, slopes exceeding 45° are particularly prone to initiating flash floods (Tariq *et al.*, 2022; Sapan *et al.*, 2023).

River density

The analysis shows that, of the eight observed watersheds, five have a dominant river-density class accounting for more than 50% of their area, as shown in Figure 5. This indicates that most watersheds have a relatively low number of rivers. However, two watersheds, Curah Nongko and Pecaron, have substantial proportions within river-density levels of 0–1, while the Wonoboyo watershed is dominated by values of 0–2. This suggests that some watersheds have low river density across a large proportion of their area, which can affect flash-flood conditions.

The upstream areas have high river density, indicated by darker colours, while the downstream areas exhibit lighter colours, indicating low river density (Fig. 5). Low river density suggests increased susceptibility to floods, consistent with findings by other researchers (Hagos *et al.*, 2022). High river density indicates that watersheds can generate large discharges (Khalifa *et al.*, 2023; Shawky and Hassan, 2023). Consequently, during extreme rainfall, the large water supply from upstream areas leads to significant water accumulation downstream, increasing the risk of flash floods.

Flash floods often occur in watersheds with low river-density values (<2), as observed in these eight watersheds. The smaller the river-density value, the fewer rivers there are in the watershed. The risk depends on the occurrence of extreme rainfall

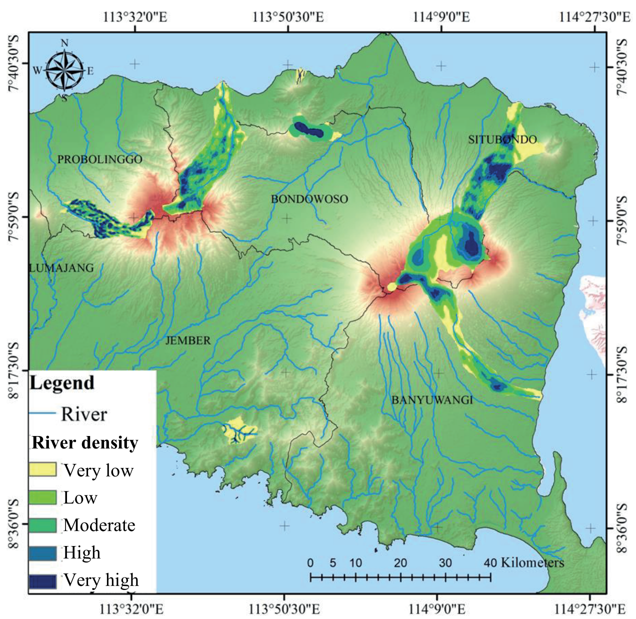

Fig. 5. River density in the eight small watersheds; source: own study

and the capacity of the rivers to contain the flow; if capacity is exceeded, flooding will occur. Additionally, flash floods in hilly upstream areas cause flood overflow to spread downstream without clear spatial confinement.

The higher the density of the water network in an area, the greater its potential impact on downstream flood-flow expansion. If heavy rainfall occurs and river capacity is insufficient, floods may spread extensively. Dense rivers tend to have better drainage systems, aiding the conveyance of excess rainwater. However, if the river network becomes too dense, high water flows can still cause flooding where drainage capacity is inadequate. High drainage density is positively correlated with flood peak and volume (Khosravi *et al.*, 2018).

TOPOGRAPHIC POSITION INDEX

The topographic position index (*TPI*) helps to identify various forms of topographic land, including different types of slopes, valleys and ridges. The *TPI* values in the eight watersheds range from −254.79 to 209.18 (Fig. 6). Negative *TPI* values depict areas lower than their surroundings, while positive *TPI* values indicate central zones higher than their neighbouring averages. The *TPI* values were divided into five classes using natural breaks. On average, the *TPI* values in the eight watersheds show a distribution approaching a bell shape. The highest area percentage is in class 3, with values close to 0, indicating relatively flat areas. The lowest area percentages are in class 1, indicating valleys, and class 5, indicating ridges.

The *TPI* factor shows strong dominance over other factors for flash floods in small watersheds across various regions. This is supported by other research on flash-flood cases (Costache, Barbulescu and Pham, 2021). However, for flood inundation in coastal or low-lying areas, its influence appears to be less significant (Youssef, Mahdi and Pourghasemi, 2023). In the case of flash floods, *TPI* characteristics in all watersheds exhibit a normal distribution pattern, indicating that watersheds with *TPI* values close to 0 (constant slope) are prone to flash floods, as shown in Figure 6. This finding is consistent with flash-flood studies elsewhere (Costache, Barbulescu and Pham, 2021; Costache *et al.*, 2021). Conversely, flood occurrences in low-lying areas are generally associated with the dominance of low *TPI* values (Choudhury *et al.*, 2022). These results indicate a difference in the dominance of *TPI* values between flash floods and flood inundation. Identifying the environmental factors influencing flash floods is crucial for mapping flash floods in small watersheds. Additionally, these watershed physical char-

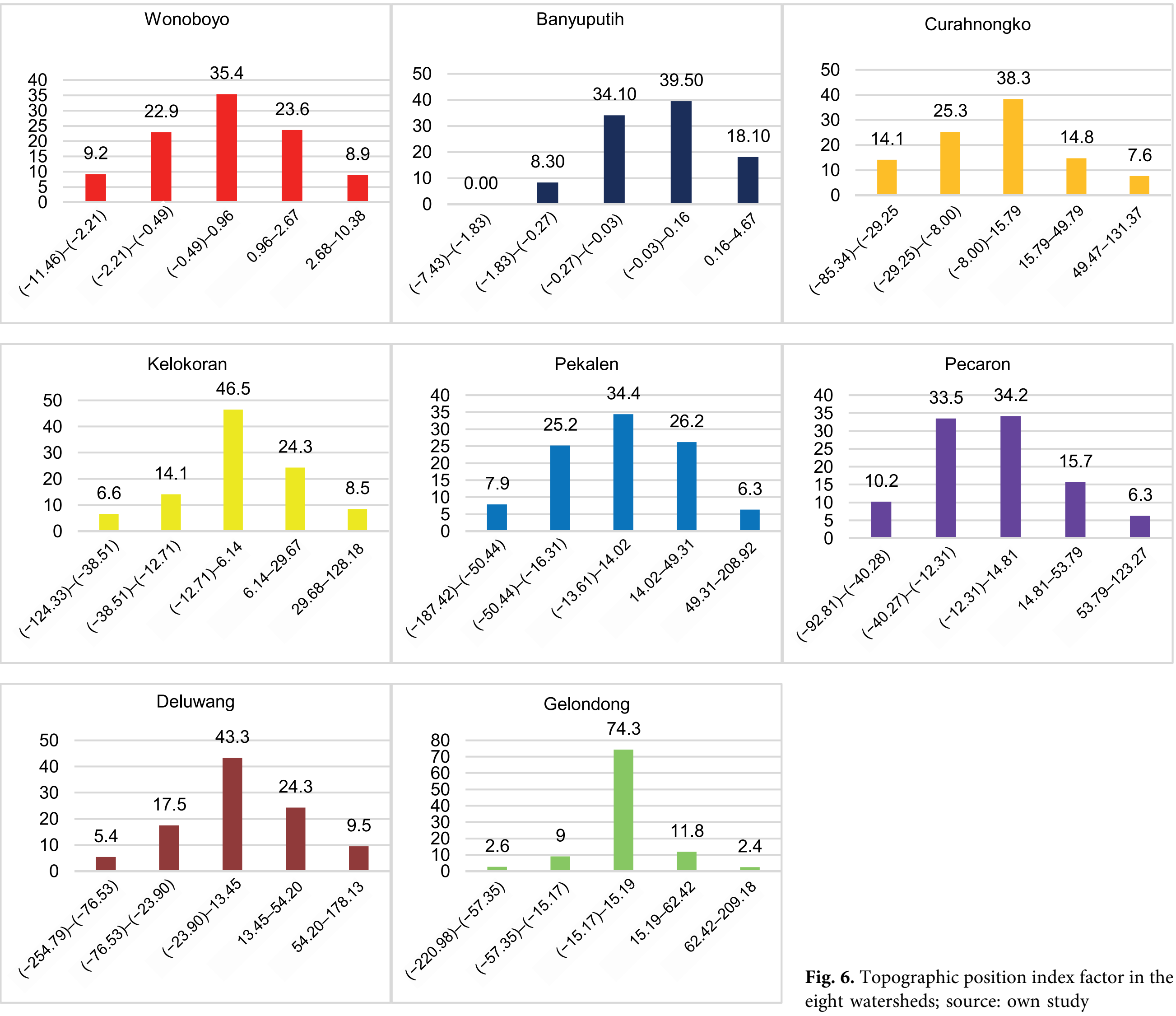

Fig. 6. Topographic position index factor in the eight watersheds; source: own study

acteristics can be used as supplementary parameters in rainfall-runoff modelling for areas prone to flash floods. The final results of this assessment can guide flood-mitigation efforts, both structural and non-structural.

CONCLUSIONS

This study demonstrates that flash-flood susceptibility in small mountainous watersheds can be effectively identified using a random forest-based spatial-modelling framework that integrates terrain, land-use and environmental variables. Model evaluation shows strong and consistent performance across eight watersheds, with training and testing receiver operating characteristic (ROC) values ranging from 0.87 to 0.98 and from 0.83 to 0.96, respectively. The results indicate that land use, slope, river density and topographic position index (*TPI*) are the most influential conditioning factors, highlighting the combined role of land-development patterns and geomorphological controls in flash-flood generation.

From a practical perspective, watersheds characterised by steep slopes, extensive land-use conversion, reduced forest cover and low *TPI* values should be prioritised for flash-flood risk management. The findings support the application of integrated mitigation strategies, including forest conservation and reforestation in upper catchments, land-use regulation, river-corridor protection and the strategic implementation of nature-based and engineered solutions. These susceptibility-based insights provide valuable guidance for watershed zoning, spatial planning and disaster risk reduction in data-limited regions.

Several limitations should be acknowledged. Uncertainties remain in the completeness and spatial accuracy of the flash-flood inventory, and land-use and normalised difference vegetation index (*NDVI*) products may be affected by classification errors. High-resolution rainfall dynamics were not explicitly incorporated due to data constraints, and the analysis focuses on susceptibility mapping rather than event-based flood prediction. Despite these limitations, the proposed framework offers a robust and transferable approach for identifying flash-flood-prone areas and supporting evidence-based mitigation planning. These recommendations are intended to support strategic planning and prioritisation rather than site-specific engineering design.

SUPPLEMENTARY MATERIAL

Supplementary material to this article can be found online at: https://www.jwld.pl/files/Supplementary_material_69_Hidayah.pdf.

CONFLICT OF INTERESTS

All authors declare that they have no conflict of interests.

REFERENCES

Abdelkebir, B., Mokhtari, E. and Engel, B. (2024) "Assessment of land use and land cover changes on hydrological responses in the Wadi Soummam Watershed, Algeria using the HEC–HMS Model," *Water Practice & Technology*, 19(9), pp. 3555–3577. Available at: https://doi.org/10.2166/wpt.2024.224.

Abedi, R. *et al.* (2022) "Flash-flood susceptibility mapping based on XGBoost, Random Forest and Boosted Regression Trees," *Geocarto International*, 37(19), pp. 5479–5496. Available at: https://doi.org/10.1080/10106049.2021.1920636.

Abid, S.K. *et al.* (2021) "Flood vulnerability and resilience: Exploring the factors that influence flooding in Sarawak," *IOP Conference Series: Earth and Environmental Science*, 802, 012059. Available at: https://doi.org/10.1088/1755-1315/802/1/012059.

Ahmad, T. and Aziz, M.N. (2019) "Data preprocessing and feature selection for Machine Learning Intrusion Detection Systems," *ICIC Express Letters*, 13(2), pp. 93–101. Available at: https://doi.org/10.24507/icicel.13.02.93.

Arora, N. and Kaur, P.D. (2020) "A Bolasso based consistent feature selection enabled Random Forest classification algorithm: An application to credit risk assessment," *Applied Soft Computing Journal*, 86, 105936. Available at: https://doi.org/10.1016/j.asoc.2019.105936.

Bathurst, J. *et al.* (2020) "Forests and floods: Using field evidence to reconcile analysis methods," *Hydrological Processes*, 34, pp. 3295–3310. Available at: https://doi.org/10.1002/hyp.13802.

BMKG (2023) *Buletin Meteorologi Stasiun Meteorologi AEK Gondang.* Edisi Bulan Agustus 2023. Badan Meteorologi Klimatologi dan Geofisika. Available at: https://stamet-aekgodang.bmkg.go.id/buletin/buletinagustus2023.pdf (Accessed: August 14, 2025).

Cao, Y. *et al.* (2020) "Flash flood susceptibility assessment based on Geodetector, certainty factor, and logistic regression analyses in Fujian Province, China," *ISPRS International Journal of Geo-Information*, 9(12), 748. Available at: https://doi.org/10.3390/ijgi9120748.

Chen, Y. *et al.* (2023) "Modeling rules of regional flash flood susceptibility prediction using different machine learning models," *Frontiers in Earth Science*,11, 1117004. Available at: https://doi.org/10.3389/feart.2023.1117004.

Choi, J.H. *et al.* (2020) "Analysis of the topography change using the RAMMS Model and numerical modeling in the Korean East Sea Estuary," *Journal of Coastal Research*, 95(sp1), pp. 422–427. Available at: https://doi.org/10.2112/SI95-082.1.

Choudhury, S. *et al.* (2022) "Flash flood susceptibility mapping using GIS-based AHP method," in B. Pradhan *et al.* (eds.) *Spatial modelling of flood risk and flood hazards. Societal implications*, pp. 119–142. Available at: https://doi.org/10.1007/978-3-030-94544-2_8.

Chowdhury, S. (2023) "MethodsX modelling hydrological factors from DEM using GIS." *MethodsX*, 10, 102062. Available at: https://doi.org/10.1016/j.mex.2023.102062.

Costache, R., Barbulescu, A. and Pham, Q.B. (2021) "Integrated framework for detecting the areas prone to flooding generated by flash-floods in small river catchments," *Water*, 13(6), 758. Available at: https://doi.org/10.3390/w13060758.

Costache, R. *et al.* (2020) "Using GIS, remote sensing, and machine learning to highlight the correlation between the land-use/land-cover changes and flash-flood potential," *Remote Sensing*, 12(9), 1422. Available at: https://doi.org/10.3390/rs12091422.

Costache, R. *et al.* (2021) "Flash-flood potential mapping using deep learning, alternating decision trees and data provided by remote sensing sensors," *Sensors*, 21(1), 280. Available at: https://doi.org/10.3390/s21010280.

Du, J., Fan, Z.-J. and Pu, J. (2020) "Comparative study on flash flood hazard assessment for Nam Ou River Basin, Lao PDR," *Natural Hazards*, 102(3), pp. 1393–1417. Available at: https://doi.org/10.1007/s11069-020-03972-3.

Hagos, Y.G. *et al.* (2022) "Flood hazard assessment and mapping using GIS integrated with multi-criteria decision analysis in Upper Awash River Basin, Ethiopia," *Applied Water Science*, 12(7), 148. Available at: https://doi.org/10.1007/s13201-022-01674-8.

Herath, P. *et al.* (2025) "A systematic review of forest cover for catchment-scale flood mitigation," *Journal of Flood Risk Management CRITICAL*, 18, e70125. Available at: https://doi.org/10.1111/jfr3.70125.

Hidayah, E., Halik, G. and Widiarti, W.Y. (2023) "Assessment of the conditioning factor for flash flood susceptibility potential based on bivariate statistical approach in the Wonoboyo Watershed in East Java, Indonesia" in S.A. Kristiawan *et al.* (eds.) *Proceedings of the 5th International Conference on Rehabilitation and Maintenance in Civil Engineering. ICRMCE 2021. Lecture Notes in Civil Engineering*, 225, Surakarta, Indonesia 8–9July 2023. Singapore: Springer. Available at: https://doi.org/10.1007/978-981-16-9348-9_49.

Hocini, N. *et al.* (2021) "Performance of automated methods for flash flood inundation mapping: A comparison of a Digital Terrain Model (DTM) filling and two hydrodynamic methods," *Hydrology and Earth System Sciences*, 25(6), pp. 2979–2995. Available at: https://doi.org/10.5194/hess-25-2979-2021.

Hu, S., Fan, Y. and Zhang, T. (2020) "Assessing the effect of land use change on surface runoff in a rapidly urbanized city: A case study of central area of Beijing," *Land*, 9(1), 17. Available at: https://doi.org/10.3390/land9010017.

Islam, A.R.M.T. *et al.* (2021) "Flood susceptibility modelling using advanced ensemble machine learning models," *Geoscience Frontiers*, 12(3), 101075. Available at: https://doi.org/10.1016/j.gsf.2020.09.006.

Kamarudin, A.N., Cox, T. and Kolamunnage-Dona, R. (2017) "Time-dependent ROC curve analysis in medical research: Current methods and applications," *BMC Medical Research Methodology*, 17(1), 53. Available at: https://doi.org/10.1186/s12874-017-0332-6.

Khalifa, A. *et al.* (2023) "Remotely sensed data, morpho-metric analysis, and integrated method approach for flood risk assessment: Case study of Wadi Al-Arish landscape, Sinai, Egypt," *Water*, 15(9), 1797. Available at: https://doi.org/10.3390/w15091797.

Khosravi, K. *et al.* (2018) "A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz Watershed, Northern Iran," *Science of the Total Environment*, 627, pp. 744–755. Available at: https://doi.org/10.1016/j.scitotenv.2018.01.266.

Kowalczyk, A. *et al.* (2023) "A multicriteria approach to different land use scenarios in the Western Carpathians with the SWAT model a multicriteria approach to different land use scenarios," *Journal of Water and Land Development*, 57, pp. 130–139. Available at: https://doi.org/10.24425/jwld.2023.145343.

Mansida, A. *et al.* (2025) "The impact of land use change on improving surface runoff , peak flood discharge, and sedimentation in the Maros Watershed," *Engineering, Technology & Applied Science Research*, 15(4), pp. 24875–2484. Available at: https://doi.org/10.48084/etasr.10826.

Mirzaei, S. *et al.* (2021) "Flood susceptibility assessment using extreme gradient boosting (EGB), Iran," *Earth Science Informatics*, 14(1), pp. 51–67. Available at: https://doi.org/10.1007/s12145-020-00530-0.

Nguyen, V.N. *et al.* (2020) "A new modeling approach for spatial prediction of flash flood with biogeography optimized CHAID tree ensemble and remote sensing data," *Remote Sensing*, 12(9), 1373. Available at: https://doi.org/10.3390/RS12091373.

Ogden, F. *et al.* (2011) "Relative importance of impervious area, drainage density, width function, and subsurface storm drainage on flood runoff from an urbanized catchment," *Water Resources Research*, 47, W12503. Available at: https://doi.org/10.1029/2011WR010550.

Saha, S., Gayen, A. and Bayen, B. (2022) "Deep learning algorithms to develop flood susceptibility map in data-scarce and ungauged river basin in India," *Stochastic Environmental Research and Risk Assessment*, 36, pp. 3295–3310. Available at: https://doi.org/10.1007/s00477-022-02195-1.

Saleh, A., Yuzir, A. and Sabtu, N. (2022) "Flash flood susceptibility mapping of sungai pinang catchment using frequency ratio," *Sains Malaysiana*, 51(1), pp. 51–65. Available at: https://doi.org/10.17576/jsm-2022-5101-05.

Sapan, E.G.A. *et al.* (2023) "Understanding flash floods in hilly tropical watersheds: A trigger factor analysis," *IOP Conference Series: Earth and Environmental Science*, 1266, 012046. Available at: https://doi.org/10.1088/1755-1315/1266/1/012046.

Sellami, E.M., Maanan, M. and Rhinane, H. (2022) "Performance of machine learning algorithms for mapping and forecasting of flash flood susceptibility in Tetouan, Morocco," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences – ISPRS Archives*, 46(4/W3-2021), pp. 305–313. Available at: https://doi.org/10.5194/isprs-archives-XLVI-4-W3-2021-305-2022.

Sharifi, E., Eitzinger, J. and Dorigo, W. (2019) "Performance of the state-of-the-art gridded precipitation products over mountainous terrain: A regional study over Austria," *Remote Sensing*, 11(17), 2018. Available at: https://doi.org/10.3390/rs11172018.

Shawky, M. and Hassan, Q.K. (2023) "Geospatial modeling based-multi-criteria decision-making for flash flood susceptibility zonation in an arid area," *Remote Sensing*, 15(10), 2561. Available at: https://doi.org/10.3390/rs15102561.

Shirzadi, A. *et al.* (2020) "A novel ensemble learning based on bayesian belief network coupled with an extreme learning machine for flash flood susceptibility mapping," *Engineering Applications of Artificial Intelligence*, 96, 103971. Available at: https://doi.org/10.1016/j.engappai.2020.103971.

Sidi, A. *et al.* (2021) "Spatiotemporal analysis of vegetation cover changes around surface water based on NDVI: A case study in Korama Basin, Southern Zinder, Niger," *Applied Water Science*, 11, 4. Available at: https://doi.org/10.1007/s13201-020-01332-x.

Sultan, D. *et al.* (2022) "Evaluation of lag time and time of concentration estimation methods in small tropical watersheds in Ethiopia," *Journal of Hydrology: Regional Studies*, 40, 101025. Available at: https://doi.org/10.1016/j.ejrh.2022.101025.

Tariq, A. *et al.* (2022) "Flash flood susceptibility assessment and zonation by integrating Analytic Hierarchy Process and Frequency Ratio Model with diverse spatial data," *Water*, 14(19), 3069. Available at: https://doi.org/10.3390/w14193069.

Wang, J.-H. *et al.* (2022) "Application of hybrid machine learning model for flood hazard zoning assessments," *Stochastic Environmental Research and Risk Assessment*, 37, pp. 395–412. Available at: https://doi.org/10.1007/s00477-022-02301-3.

Widyasasi, D. *et al.* (2024) "Land use change and Soil Conservation Services Curve Number (SCS-CN) in Karangmumus Watershed Samarinda," *International Journal of Engineering Trends and Technology*, 72(10), pp. 1–9. Available at: https://doi.org/10.14445/22315381/IJETT-V72I10P101.

Yang, H. *et al.* (2017) "The standardization and harmonization of land cover classification systems towards harmonized datasets: A review," *ISPRS International Journal of Geo-Information*, 6(5), 154. Available at: https://doi.org/10.3390/ijgi6050154.

Yao, J. *et al.* (2022) "Applications of stacking/blending ensemble learning approaches for evaluating flash flood susceptibility," *International Journal of Applied Earth Observation and Geoinformation*, 112, 102932. Available at: https://doi.org/10.1016/j.jag.2022.102932.

Youssef, A.M., Mahdi, A.M. and Pourghasemi, H.R. (2023) "Optimal flood susceptibility model based on performance comparisons of LR, EGB, and RF algorithms," *Natural Hazards*, 115, pp. 1071–1096. Available at: https://doi.org/10.1007/s11069-022-05584-5.

Yuan, W. *et al.* (2021) "Research on the critical rainfall of flash floods in small watersheds based on the design of characteristic rainfall patterns," *Water Resources Management*, 35(10), pp. 3297–3319. Available at: https://doi.org/10.1007/s11269-021-02893-5.

Zhang, Y. *et al.* (2021) "Assessment on the effectiveness of urban stormwater management," *Water*, 13(1), 4. Available at: https://doi.org/10.3390/w13010004.