Dynamic Spatial Study for Flood Risk Mitigation Using GIS and Machine Learning Approaches

(Case Study: Citarum hulu and Cirasea Sub-Watersheds)

Sheila Nanda Fadjrilia [1*], Rina Marina Masri[1], Rakhmat Yusuf[1]
[1]Civil Engineering Study Program, Indonesia University of Education
*Corresponding Author: Sheila Nanda Fadjrilia. Email: sheilanandaf21@gmail.com

Abstract

Flooding is a recurring disaster in Indonesia, particularly in the downstream areas of the Citarum Hulu and Cirasea Sub-Watersheds, which are highly vulnerable due to rapid land-use change. This study aims to analyze the impact of land cover transformation on flood vulnerability using an integrated spatial and computational approach. The research employed Geographic Information Systems (GIS) and the Random Forest algorithm to process land cover data from 2014 to 2024, alongside environmental variables such as soil type, slope, rainfall, river proximity, and land use. Hydrological simulations were conducted to assess peak discharge under various land use scenarios. The results show that land cover changes especially the expansion of built-up areas and the reduction of vegetated and agricultural land significantly increase flood risk. Machine learning classification confirmed that these land transformations are strongly correlated with historical flood-prone zones. Hydrological modeling further demonstrated that mitigation strategies such as river normalization and the construction of retention ponds can reduce peak discharge from 253 m^3/s to 169 m^3/s. These findings highlight the critical role of land use planning in flood mitigation. The integration of GIS and machine learning offers an effective, adaptive, and evidence-based framework for sustainable watershed and disaster risk management.

Keywords: Land use change, flood vulnerability, Geographic Information Systems (GIS), Random Forest, retention ponds

1. INTRODUCTION

Land-use change has become a critical environmental issue in Indonesia, particularly within the Citarum Hulu sub-watershed, which has experienced extensive land conversion due to increasing population and urban development. In the Cirasea sub-watershed, built-up areas increased by 4.74%, open land by 6.62%, shrubland by 2.08%, and plantations by 1.05%.[2]. These transformations have reduced environmental carrying capacity and heightened flood risk in downstream areas, especially Majalaya Subdistrict, which, according to BNPB, experiences flooding up to ten times per year.

The conversion of green open spaces and infiltration areas has worsened the hydrological balance, resulting in waterlogging during the rainy season and drought during the dry season[1]. Geographic Information System (GIS) technology enables a quantitative spatial analysis of land-use changes, while the Random Forest algorithm is effective in modeling the relationship between land cover and flood vulnerability. A spatial dynamic approach is essential to identify high risk zones and support risk-based adaptive planning, including flood-prone zoning and the development of infrastructure such as drainage canals and retention ponds.[4]. In addition, the integration of HEC-RAS software is used to simulate hydrodynamic flow and comprehensively evaluate the effectiveness of flood mitigation efforts in Majalaya Subdistrict.

This study aims to analyze the relationship between land-use change and increased flood vulnerability through a spatial and hydrological approach. Geographic Information System (GIS) is used to map land cover changes from 2014 to 2024 and to process physical-environmental parameters such as slope, soil type, rainfall, distance to rivers, and land-use patterns. The Random Forest algorithm is then applied to identify the most influential variables contributing to flood occurrences. The model is validated using historical flood event data and performance accuracy evaluation. The novelty of this study lies in its spatial data driven analytical approach, which quantitatively identifies the dominant factors causing flooding. Rather than merely mapping flood-prone zones, this approach offers a deeper scientific foundation for adaptive and targeted flood mitigation planning, particularly in areas facing high pressure from land-use conversion.

1.1 Land Changes in the Citarum WatershedThe Upper Citarum

The Upper Citarum Watershed (DAS Citarum Hulu) serves as a conservation area characterized by high drainage density and steep slopes (>15%). However, it has experienced rapid land use changes. The watershed's infiltration function has declined due to deforestation, the expansion of agricultural land, and the growth of built-up areas, leading to an increased risk of annual flooding, particularly in regions near the water supply flow into the Citarum River. Existing rehabilitation and spatial planning programs have not been effective in curbing the rate of environmental degradation. Simultaneous land use changes in the upstream area contribute significantly to river overflow and the rising flood potential. Therefore, assessing land use change is a crucial step in planning watershed-based regional management.[3]

1.2 The Impact of Land-Use Change on Hydrological Characteristics

Flood events in downstream areas cannot be separated from rainfall patterns in the upstream regions. Land-use changes significantly affect hydrological characteristics and alter flow behavior. Several studies have shown that the expansion of impervious surfaces due to land conversion reduces infiltration, decreases groundwater recharge, and increases surface runoff. The lowering of the groundwater table directly affects the reduction of baseflow, while conversely, increased surface runoff leads to higher discharge rates. Visible changes in watershed characteristics include a rise in flood inundation points, increased soil erosion in agricultural areas, and sediment accumulation in rivers.[3]

1.3 flood vulnerability

Floods can generally be classified into two categories: natural floods and anthropogenic floods. Natural floods are influenced by factors such as rainfall intensity, physiographic conditions, erosion and sedimentation, river capacity, drainage capacity, and tidal effects. Meanwhile, anthropogenic floods are caused by human activities that lead to environmental degradation, including changes in watershed conditions, settlements near riverbanks, damaged land drainage systems, malfunctioning flood control structures, deforestation (loss of natural vegetation), and improper planning of flood control systems[5]. Several parameters have been identified as significantly influencing flood vulnerability levels, including the following:

1) Rainfall

Regions with high rainfall are more susceptible to flooding. The greater the annual precipitation in an area, the higher its potential for flood events.

2) Slope Gradient

Slope gradient significantly influences the speed and volume of surface runoff, surface drainage, land use patterns, and soil erosion. It is generally assumed that the gentler the slope, the slower the surface runoff will be, thereby increasing the likelihood of water accumulation and flooding. Conversely, steeper slopes accelerate surface runoff, allowing rainwater to flow more quickly and reducing the chance of water ponding or flooding.

3) Soil type

Soil texture plays a critical role in determining infiltration capacity. Coarse-textured soils, such as sand, generally have high infiltration rates, allowing rainwater to be absorbed quickly. In contrast, fine-textured soils like clay and loam have low infiltration capacity and saturate more easily, thereby increasing the potential for surface runoff and flood occurrences.

4) Land Use

Land use is a major factor contributing to surface water accumulation and flooding in a region, as it directly influences the volume of surface runoff generated when rainfall exceeds the soil's infiltration capacity. Areas dominated by vegetation are less likely to experience flooding because vegetation facilitates higher infiltration rates and slows the flow of runoff, allowing more time before it reaches the river. In contrast, non-vegetated or impervious areas produce faster and larger volumes of runoff, increasing flood risk.

5) Distance to River

Distance from the river is used to determine the extent of areas influenced by river flow and the potential for flooding to spread to surrounding regions. River buffers are used as one of the key parameters in identifying flood hazard zones. Areas closer to the river are generally more vulnerable to inundation due to overflow during high discharge events.

1.4 Flood Vulnerability Level

Flood vulnerability scores are calculated by summing the weighted values of four key parameters. The flood vulnerability index is determined using the following equation[7]:

$$K = Wi \times Xi \quad (1)$$

Keterangan
K = Flood Vulnerability Index
Wi = Weight assigned to the *i-th* parameter
Xi = Score of the *i-th* parameter1.5 Banjir di hilir (Kecamatan Majalaya)

The downstream area of the Citarum River, particularly Majalaya Subdistrict, is a major industrial zone with numerous textile factories. As a result, the river has become a disposal site for industrial wastewater heavily contaminated with solid waste from factories, leading to significant sedimentation and shallowing of the riverbed. This condition is one of the primary factors contributing to river overflow during the rainy season. Consequently, residential areas along the Citarum River such as Majalaya, Bojongsoang, and Dayeuhkolot have become highly flood-prone zones.

Figure 1 Flooding in Majalaya Subdistrict
Source: artikel detikjabar

According to a news report from September 2024, heavy rainfall caused flooding up to 60 cm deep on Jalan Raya Cidawolong. Floodwaters also inundated Jalan Raya Laswi, reaching depths of 40 to 60 cm. Despite the severity, several vehicles were still able to pass through, and the floodwaters were expected to recede within 4–5 hours.

1.5 Geographic Information System (GIS)

A Geographic Information System (GIS) is a computer-based system used for mapping and analyzing various phenomena and events that occur on the Earth's surface. As a type of Computer-Based Information System (CBIS), GIS is specifically designed to store and process spatial information. It facilitates the collection, storage, and analysis of geographic objects and phenomena, with location being a fundamental and essential attribute for analysis.

1.6 Machine learning

The application of machine learning using the Random Forest algorithm is categorized as a *bagging* method. This approach generates multiple decision trees from sample data, where the creation of each tree during training is independent of the others. Final decisions are made based on majority voting. Random Forest is based on two core principles: constructing an ensemble of trees through bootstrap aggregation (*bagging with replacement*), and randomly selecting features for each tree to improve generalization and reduce overfitting.

2. METHOD

This study employed a quantitative descriptive method to examine the spatial dynamics of land cover changes and their relationship to flood vulnerability. The objective was to provide a systematic and accurate description of the conditions in the Upper Citarum and Cirasea Sub-watersheds and their impact on flooding in the Majalaya Subdistrict.

2.1 Research Object

The research object in this study is the Citarum hulu and Cirasea Sub-watersheds, which hydrologically form part of the upstream region of the Citarum Watershed. These areas drain into Majalaya Subdistrict, Bandung Regency a densely populated downstream area. The sub-watersheds are experiencing increasing development pressure, including expansion of built-up land, reduction of forested areas, and the conversion of open land into settlements and agricultural zones. These changes have led to a high frequency of annual flooding, particularly in zones adjacent to the riverbanks and riparian areas. Therefore, the analysis focuses on the spatial and temporal relationships between land cover changes and flood intensity in the downstream region.

2.2 Research Stage

This research was conducted through three main stages:

1) Land Use Change Analysis: Using ArcGIS software, overlay analysis was performed on land cover data from 2014 and 2024 to identify spatial changes within the sub-watershed area.
2) Flood Factor Modeling: Five physical parameters soil type, slope gradient, rainfall, distance to river, and land cover were input into a Random Forest model to determine their respective contributions to flood occurrences.
3) Flood Vulnerability Mapping and Mitigation: The modeling results were converted into flood vulnerability zoning maps, which served as the basis for developing spatially-based mitigation recommendations.

2.3 Data Inventory

The data used in this study include spatial, climatological, and hydrological datasets. All data were obtained from official institutions and processed using GIS and machine learning approaches. Table 6 summarizes the types and sources of data

Table 1 Data Source

No	Primary Data Type	Data Source	Year
1	Subdistrict Boundary Data	Seluruh Indonesia Lengkap - Lapak GIS)	2019
2	Rainfall Data	NASA – Giovanni/BINTEK Sumber daya air	2014-2024
3	Soil Type Data	WARUNGMAPPING.MY.ID (Data shp)	2020
4	River Network Data	Balai besar Wilayah sungai Citarum	2024
5	Flood History	BPBD Kab. Bandung	2024
6	DEMNAS	Badan Informasi Geospasial (https://tanahair.indonesia.go.id/)	2024
7	WLR Majalaya Data	Balai besar wilayah sungai Citarum	2024

3 RESULTS AND DISCUSSION

3.1 Land Use Change in the Cirasea and Citarum hulu Sub-Watersheds

Table 2 Land use change

No	Jenis Tutupan Lahan	2014 (km^2)	2015 (km^2)	2016 (km^2)	2017 (km^2)	2018 (km^2)
1	Forest	110,120117	105,22	100,2	98,43	96,31
2	Dry Land / Barren Land	4,484072	4,48	4,08	4,9026	3,88
3	Built-up Area	23,916242	26,272312	28,08	29,16	30,24
4	Plantation	9,6834	11,68	10,135243	12,0001	14,9753
5	Agricultural Land	99,987	104,757654	106,227654	100,3192	102,0444
6	Rice Field / Paddy Field	63,412686	59,198638	62,55	66,45	63,39
7	Water Body	6,2222364	6,222556	6,56	6,57	6,99
	Jumlah	317,83	317,83	317,83	317,83	317,83

Continuation of the table 4.2

2019 (km^2)	2020 (km^2)	2021 (km^2)	2022 (km^2)	2023 (km^2)	2024 (km^2)
93,24	90,26	86,95	83,6453	78,78	77
4,85	4,346104	4,346104	4,024	6,234	5,149
31,32	37,4	43,48	44,39	49,25	53,265
15,709	19,03	18,56171	19,56171	20,56171	22,713
111,5302	107,0347	111,0347	117,0347	115,0347	111,179
54,08	52,75	46,367	41,97499	40,65499	41,127
7,1	7,01	7,091	7,199	7,315	7,5003
317,83	317,83	317,83	317,83	317,83	317,83

The main findings indicate that built-up areas increased by 29.4%, with the most significant expansion occurring in the Majalaya region and along primary transportation corridors. Meanwhile, rice fields and agricultural land decreased by 18.6%, particularly in low-lying zones and areas adjacent to riverbanks. Dense vegetation cover also declined. These trends reflect a shift in land utilization—from agricultural and green zones to residential and economic development areas. Spatial planning should address two essential aspects:[6]

1. Residential Development, including the provision of housing and supporting infrastructure.
2. Agricultural and Plantation Development, which are vital for food security and environmental sustainability.

However, the study reveals that the conversion of agricultural land into residential areas has occurred without spatial planning based on land carrying capacity. This has resulted in:

1. A reduction in water infiltration capacity, contributing to increased flood risk.
2. Conflicts in land-use interests, especially between economic demands (settlements and industries) and water resource conservation.
3. In other words, although housing development is a legitimate public need, neglecting spatial suitability based on ecological function can lead to environmental degradation and elevate the risk of disasters.

3.2 Machine Learning

This model was implemented using five spatial input variables, namely:

1. Land use
2. Slope gradient
3. Average annual rainfall
4. Distance to river
5. Soil type

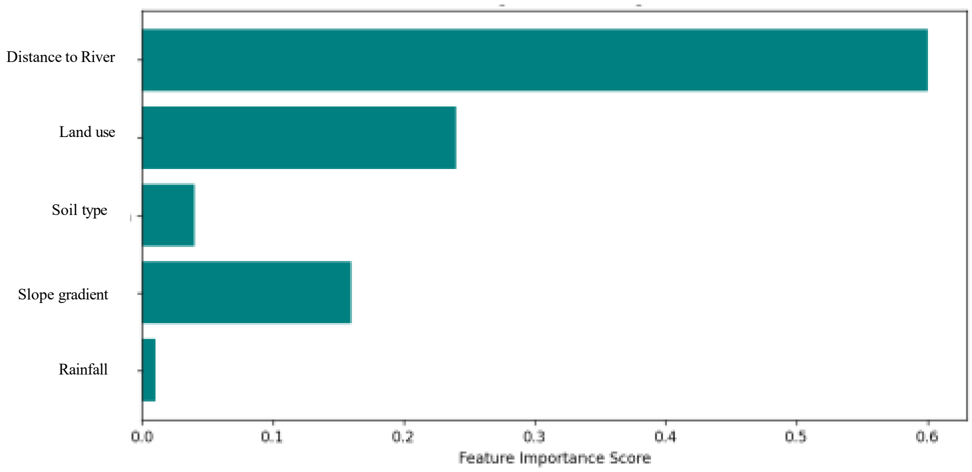

Figure 2 Graph of Flood Influencing Factors

Based on the variable importance graph, distance to river had the greatest influence on flood vulnerability, with a score exceeding 0.6, indicating that areas located near rivers are more prone to inundation during high-discharge events. In contrast, land use/land cover (LULC) initially showed a moderate contribution (approximately 0.15), reflecting the impact of land conversion on flood risk. However, when the model was retrained using multi-temporal data (2014–2024), LULC became the most dominant factor, surpassing distance to river. This shift highlights that changes in land cover—particularly the conversion of vegetated areas into built-up or open landplay a critical role in increasing flood vulnerability due to reduced infiltration capacity and higher surface runoff.

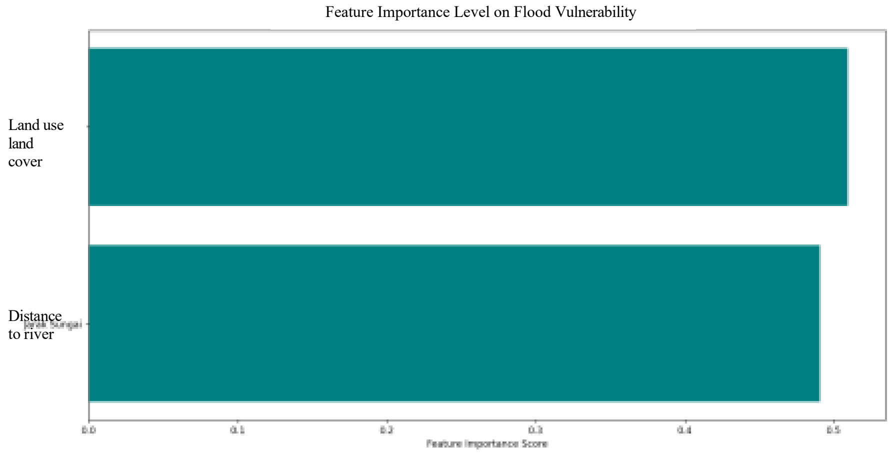

Figure 3 Graphs Showing the Impact of Floods over Temporal Changes

Unlike distance to river, which is a static feature, land use/land cover (LULC) is dynamic and changes from year to year. This temporal variation enhances the responsiveness of the model by providing updated information that reflects evolving flood risk. In the Random Forest modeling process, the importance weight of LULC increased when multi-temporal data were applied, indicating that land cover change such as the conversion from vegetation to settlements has a greater influence on flood risk.

While distance to river remains hydrologically significant, its contribution to the model declined due to its constant nature over time. Moreover, findings from the village of Biru show that flooding also occurred in areas far from the river, primarily due to land conversion and poor local drainage. This demonstrates that flooding in Majalaya is not solely dependent on river proximity, but is also strongly affected by land use planning and the adequacy of drainage infrastructure.

3.3 Land Use Changes and Flooding in Majalaya

Based on hydrological analysis using the Curve Number (CN) method from 2014 to 2024, there is a consistent upward trend in CN values from 50.55 in 2014 to 52.27 in 2024. This increase reflects a decline in land infiltration capacity, primarily due to the expansion of built-up areas and the reduction of natural vegetation cover in the Cirasea Sub-watershed area.

As the CN values increased, the surface runoff (Q) also rose. In 2014, the runoff was recorded at 173 m^3/s, while by 2024 it had reached 253 m^3/s. This trend illustrates a continuous rise in the volume of rainfall converted into surface runoff, which directly contributes to the increasing flood potential in downstream areas—particularly in Majalaya Subdistrict.

3.4 Flood Vulnerability Mapping in the Downstream Area (Majalaya Subdistrict)

The flood vulnerability map was generated through spatial analysis of five key physical parameters that significantly contribute to determining a region's susceptibility to flooding. These parameters include: Soil type, Slope gradient, Land cover, distance to river, rainfall intensity

Table 3 Parameter Scores per Village

Village	Rainfall	Slope	Soil Type	Land Use	Distance to River
Neglasari	3	2,30	2	3,26	0,95
Wangisagara	3	2,01	2	3,01	2,30
Padamulya	3	2,31	2	4,00	0,51
Sukamukti	3	2,39	2	3,47	0,01
Padaulun	3	2,29	2	3,41	1,43
Biru	3	2,88	2	4,60	0,99
Sukamaju	3	2,30	4	3,66	2,44
Majasetra	3	2,40	4	3,68	2,11
Majalaya	3	2,12	2	4,17	1,80
Majakerta	3	2,12	4	4,12	0,60
Bojong	3	2,21	2	4,30	4,00

Table 4 Flood Vulnerability Parameter Weights

No	Parameter	Weight
1	Jenis Tanah	0,1
2	Kemiringan lereng	0,25
3	Tutupan lahan	0,25
4	Curah hujan	0,15
5	Jarak ke sungai	0,25

Table 5 lood Vulnerability Classification

No	Skor	Kelas kerawanan
1	<2	Safe
2	2-2,5	Moderately Vulnerable
3	2,5-3	Vulnerable
4	>3	Highly Vulnerable

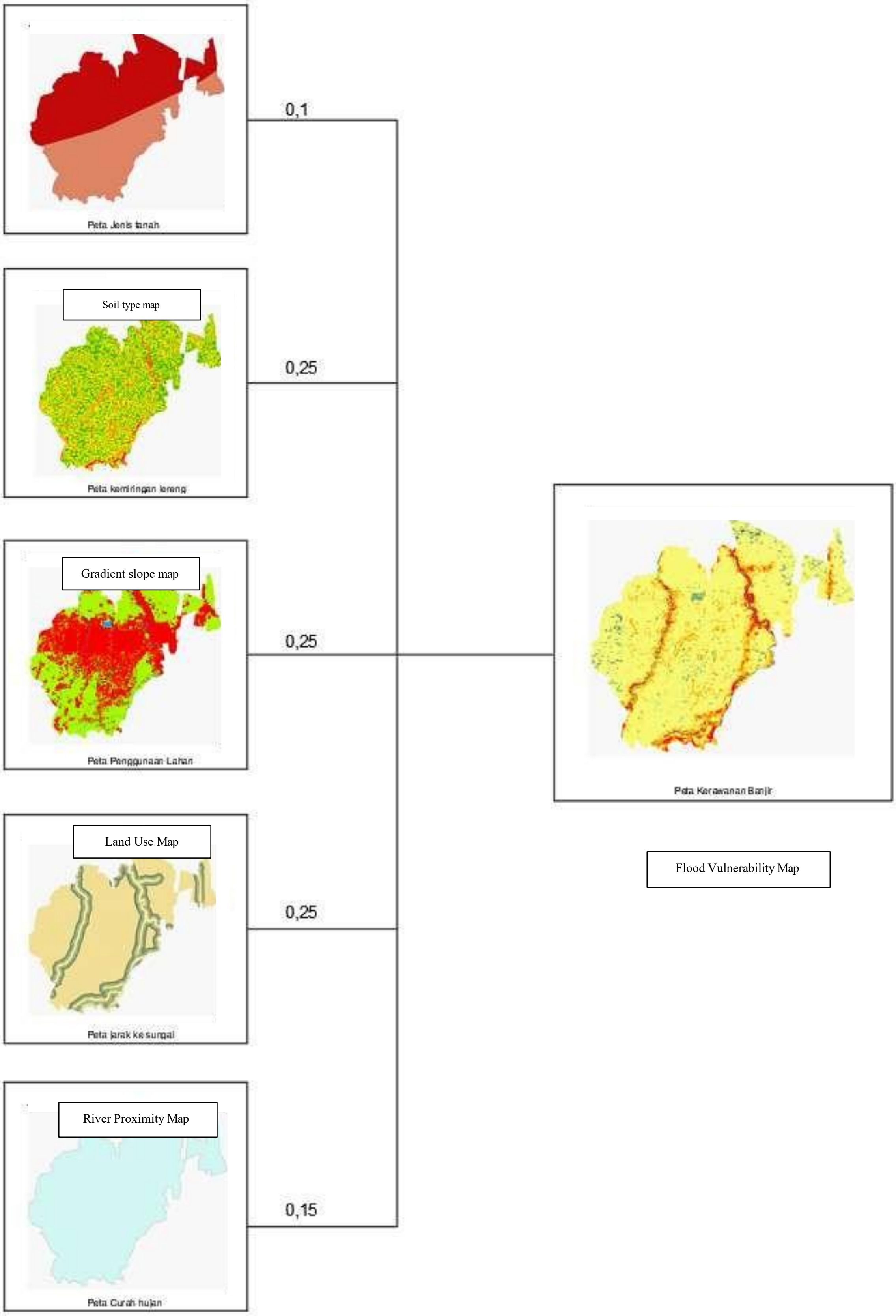

Figure 2 Attribute Decision Tree

Rainfall Map

Graphs Showing the Impact of Floods over Temporal Changes

Figure 4 Decision Tree Attributes

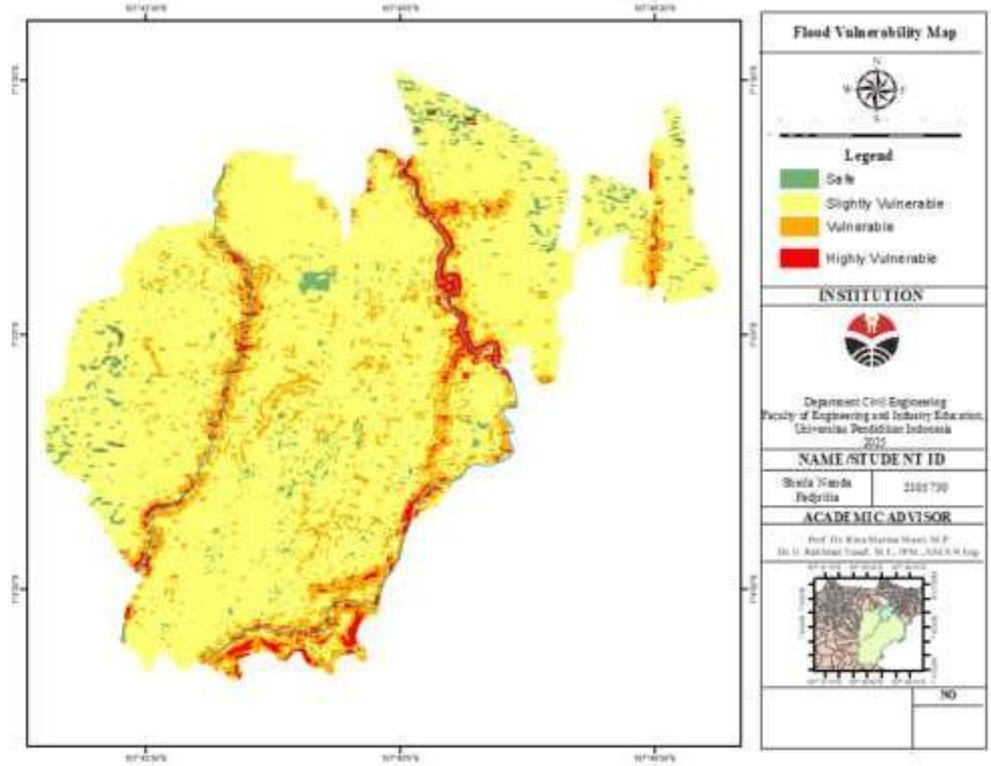

Figure 5 Flood Vulnerability Map

The flood vulnerability map for Majalaya Subdistrict was produced through a spatial overlay of five physical parameters: soil type, slope gradient, land use/land cover, distance to river, and rainfall. Each parameter was weighted according to its contribution to flood risk. The resulting map (Figure 7) classifies the area into four levels of flood vulnerability: safe, moderately vulnerable, vulnerable, and highly vulnerable.

The highly vulnerable zones are generally located along the main river channels, characterized by easily saturated latosol soils, flat slopes, land use dominated by rice fields and residential areas, and close proximity to rivers. The vulnerable class is distributed in areas at a moderate distance from the river and is primarily composed of open land or rice fields.

In contrast, the moderately vulnerable to safe zones are mostly found in the western and southwestern parts of the subdistrict, where the topography is higher, slopes are steeper, and natural vegetation provides effective infiltration zones.

3.5 Mitigation Recommendations

1) River Normalization
River normalization is a technical water resource management strategy aimed at increasing river flow capacity, particularly during peak discharge events caused by heavy rainfall. This strategy includes sediment dredging, cross-sectional widening, base slope correction, and riverbank reinforcement.

In Majalaya Subdistrict, normalization is a priority due to riverbed sedimentation and channel narrowing caused by residential and agricultural activities. These conditions reduce river storage capacity and heighten runoff risk.

Recommended actions include mapping the current river cross-sections, analyzing the design discharge (Q_{50}), adjusting channel dimensions (width, depth, slope), and constructing embankments where necessary. These efforts are expected to enhance hydraulic capacity and significantly reduce flood risk in downstream areas.

2) Retention Pond
As an adaptive structural strategy, the construction of retention ponds is recommended to manage stormwater runoff in the downstream area of the Majalaya sub-watershed. These ponds temporarily store peak discharge before gradually releasing it into the river, thus alleviating flow pressure during extreme rainfall events. Based on

simulations using the SCS Unit Hydrograph method, the peak discharge of 253 m³/s exceeds the capacity of the existing river channel, necessitating peak shaving through temporary storage.

Retention ponds are favored over levees due to their superior hydrological effectiveness, lower failure risk, and added benefit of contributing to green open space. A study by Susilo et al. (2021) showed that retention ponds can reduce peak discharge by up to 45% and extend the time of concentration by 25%, making them a multifunctional solution for urban flood control. In this analysis, retention ponds were shown to reduce peak discharge by 33.20%, especially when combined with river normalization.

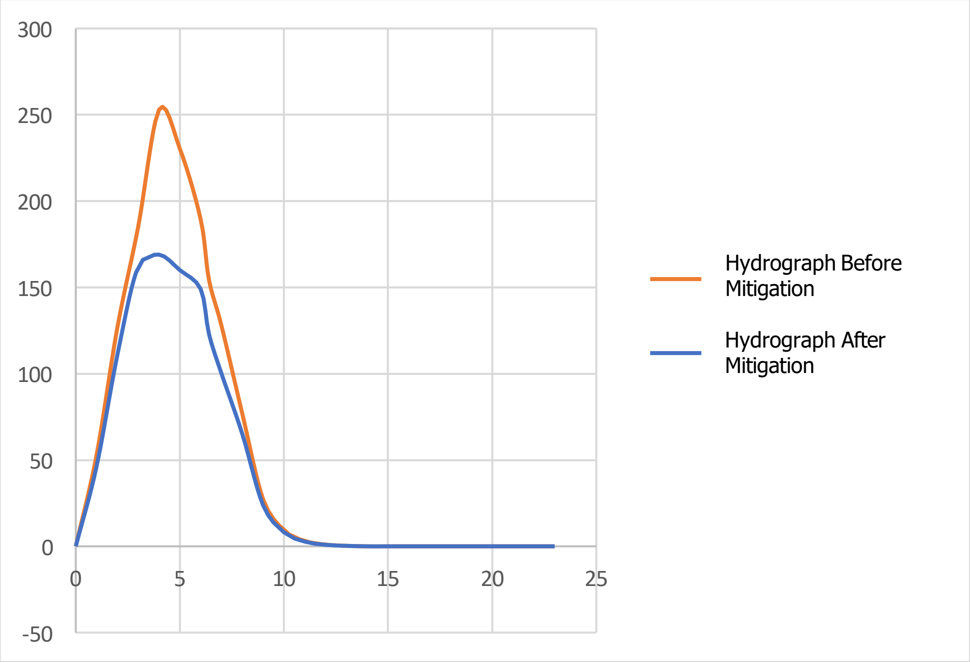

Figure 3 Flood Hydrograph

The figure illustrates the change in hydrograph patterns before and after mitigation efforts, highlighting the reduction in peak discharge and the delay in peak timing resulting from the implementation of river normalization and retention pond strategies.

4. CONCLUSIONS

1) During the 2014–2024 period, the Cirasea and Citarum hulu sub-watersheds experienced significant land use changes. There was a marked decrease in forest and shrubland areas, accompanied by an increase in built-up and open lands. These changes have led to a reduction in infiltration capacity and an increase in surface runoff. Spatial overlay analysis revealed that areas with high rates of land conversion tend to exhibit a higher frequency of flood events, particularly in the downstream region of Majalaya Subdistrict.
2) A Machine learning analysis identified distance to river and land use/land cover (LULC) as the most influential variables affecting flood vulnerability. In the initial model, distance to river emerged as the dominant factor with an importance score exceeding 0.6. However, after incorporating multi-temporal LULC data and annual flood occurrences, the role of land cover significantly increased, indicating that land use change is the primary determinant of rising flood risk in the study area. Other factors such as slope and rainfall also contribute, albeit with relatively lower influence.
3) In response to these conditions, the recommended mitigation strategies include river normalization and the construction of retention ponds. River normalization aims to restore the river's flow capacity through dredging, widening, and bank reinforcement. Meanwhile, retention ponds serve as temporary storage for stormwater runoff, reducing peak discharge loads. Hydrological simulations demonstrate that retention ponds can substantially reduce peak flow and provide delayed runoff timing. This approach is considered more effective, safer, and multifunctional compared to permanent levee structures.

ACKNOWLEDGMENT

The author would like to express sincere gratitude to all parties who have provided support and guidance throughout the preparation of this research. Special appreciation is extended to Prof. Dr. Rina Marina Masri, M.P. and Drs. Ir. Rakhmat Yusuf, M.T., IPM., ASEAN Eng., as academic supervisors. Their continuous guidance, constructive feedback, and invaluable academic insights have played a crucial role in shaping the direction and quality of this research.

REFERENCES

1. Christian, Y., Asdak, C., & Kendarto, D. R. (2021). Analisis Perubahan Penggunaan Lahan di Kabupaten Bandung Barat. *Jurnal Teknotan*, *15*(1), 15.
2. Davik R, Rachman LM, Hidayat Y, Ridwansyah I. Dinamika perubahan penggunaan lahan di Sub-DAS Cirasea (DAS Citarum Hulu). Jurnal Penelitian Pengelolaan Daerah Aliran Sungai. 2022;6(2):123–33.
3. Fadhil, M. Y., Hidayat, Y., Murtilaksono, K., & Baskoro, D. P. T. (2021). Perubahan penggunaan lahan dan karakteristik hidrologi DAS Citarum Hulu. *Jurnal Ilmu Pertanian Indonesia*, *26*(2), 213-220.
4. Fairuzzen, M. R., Merpaung, V. H., Putra, A. A., & Malik, A. A. (2024). Peran Tata Ruang dalam Mitigasi Risiko Pembangunan dan Pencegahan Bencana Alam. *Interdisciplinary Explorations in Research Journal*, *2*(3), 1497-1516.
5. Ka'u, A. A., Takumansang, E. D., & Sembel, A. (2021). Analisis Tingkat Kerawanan Banjir di Kecamatan Sangtombolang Kabupaten Bolaang Mongondow. *Spasial*, *8*(3), 291-302.
6. Rachman M, Putra A, Adi D. Pemanfaatan ruang dan dampaknya terhadap kelestarian alam. Jurnal Tata Ruang dan Pembangunan. 2023;5(1):33–42.
7. Seprianto, M., Anggo, M., Harudu, L., & Aldiansyah, S. (2024). Pemetaan Daerah Potensi Rawan Banjir Menggunakan Metode Overlay. *Jurnal Penelitian Pendidikan Geografi*, *9*(4), 214-226.