Predicting Flood and Landslide Hazards in Indonesia:
A Machine Learning Approach for Risk Assessment and Urban Planning

Sujung Heo[1], Dong Kun Lee [1,2]

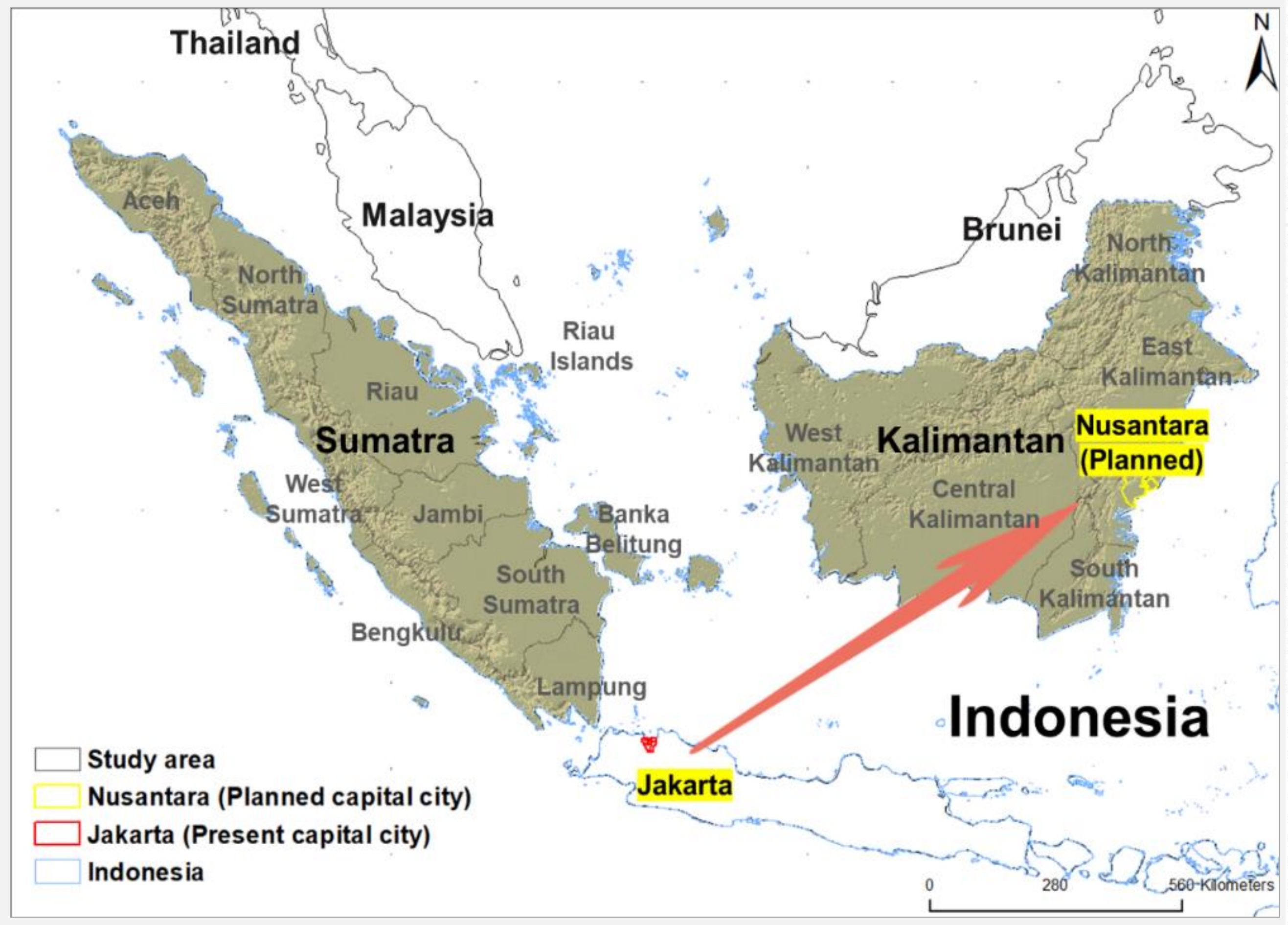
[1]Interdisciplinary Program and Life Science, Seoul National University, Seoul, Rep. of Korea (christina_h@snu.ac.kr), [2]Department of Landscape Architecture and Rural System Engineering, Seoul National University, Seoul, Rep. of Korea

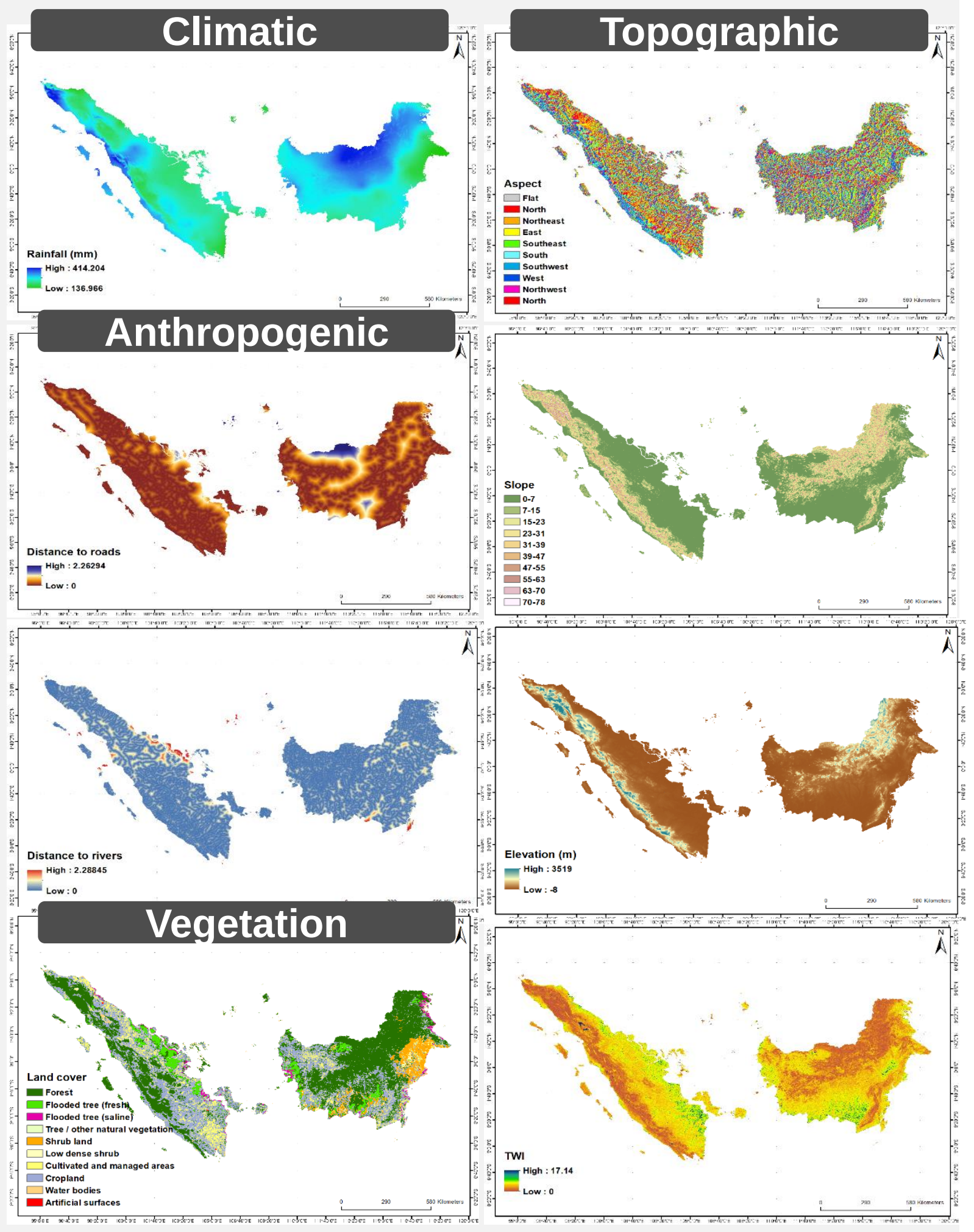
INTRODUCTION

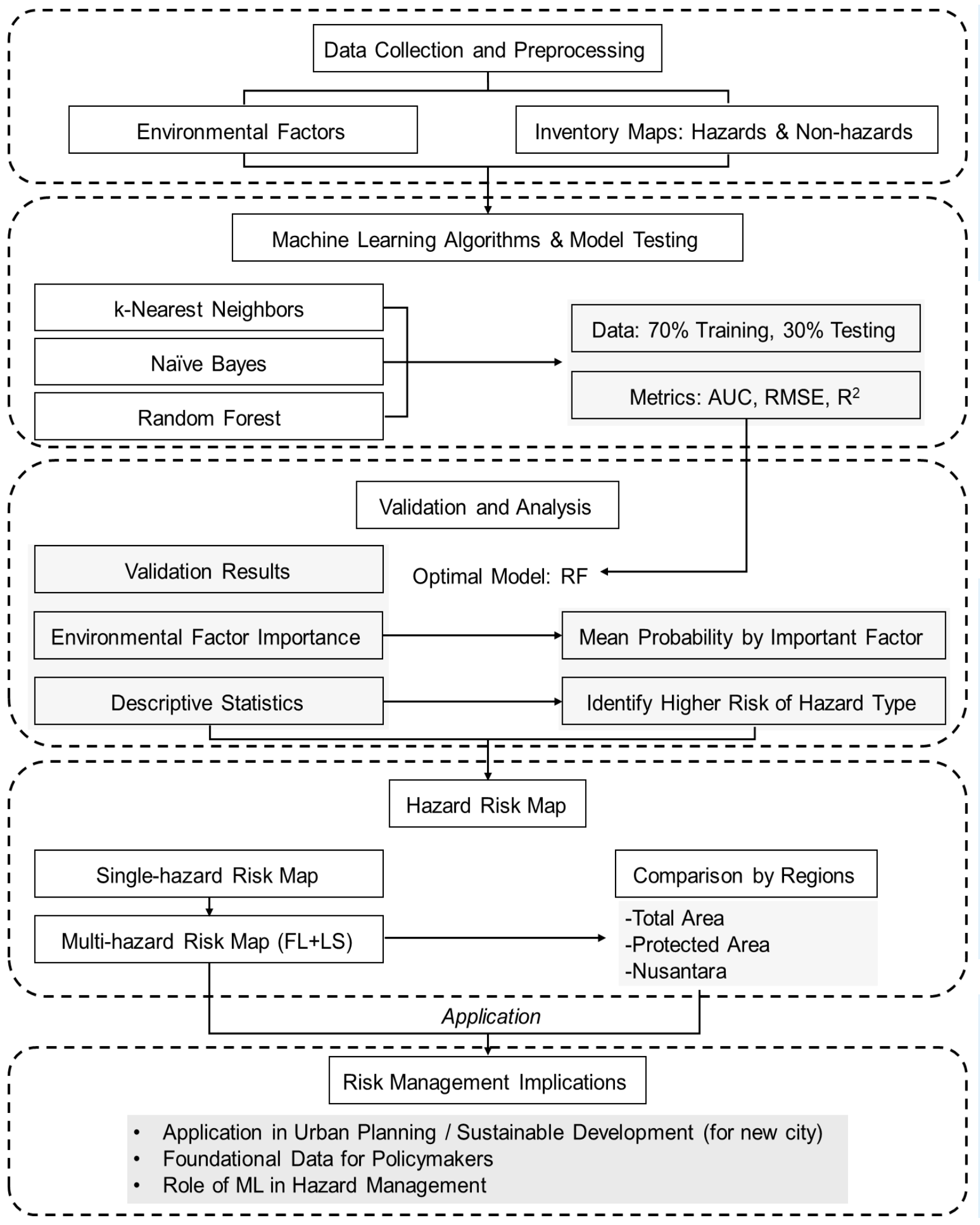
- Planning homes in disaster-prone areas should involve a comprehensive multi-hazard analysis, as per Agenda 21, and thorough disaster management covering prevention, mitigation, preparedness, response, and recovery, as recommended by the UN's Action Plan for Johannesburg (2002).
- To reduce hazard risk, it is crucial to create assessment techniques, which are advantageous to share with other stakeholders, such as the business and local government sectors. The target area's susceptibility to floods and landslides is evaluated in this study's multi-hazard.
- This study presents a comprehensive multi-hazard assessment for the major islands of Indonesia, focusing on the interactions between floods and landslides. These are the research issues:

1) How do various environmental variables interact to influence the occurrence of floods and landslides in Kalimantan and Sumatra?
2) How do different machine learning models compare in terms of accuracy and computational efficiency when applied to multi-hazard assessments?
3) What strategies can be used to integrate multi-hazard risk maps into urban planning and infrastructure development to enhance disaster resilience in the new capital region and other vulnerable areas?

MATERIALS & METHODS







Data Collection and Preprocessing
Environmental Factors
Inventory Maps: Hazards & Non-hazards
Machine Learning Algorithms & Model Testing
k-Nearest Neighbors
Naïve Bayes
Random Forest
Data: 70% Training, 30% Testing
Metrics: AUC, RMSE, R^2
Validation and Analysis
Validation Results
Optimal Model: RF
Environmental Factor Importance
Mean Probability by Important Factor
Descriptive Statistics
Identify Higher Risk of Hazard Type
Hazard Risk Map
Single-hazard Risk Map
Multi-hazard Risk Map (FL+LS)
Comparison by Regions
-Total Area
-Protected Area
-Nusantara
Application
Risk Management Implications

- Application in Urban Planning / Sustainable Development (for new city)
- Foundational Data for Policymakers
- Role of ML in Hazard Management

- **NB**
  - A stochastic statistical approach founded on Bayes' theorem, which utilizes to calculate the posterior probability.

$$P(A \mid B) = \frac{P(A) \cdot (B \mid A)}{P(B)}$$

Where P(A) and P(B) represent the prior probability, P(B|A) represents the likelihood, and P(A|B) represents the posterior probability.

- **kNN**
  - Strength in handling non-linear data distributions. The value of k is usually chosen between 3 to 10, as too low a value can lead to overfitting, and too high a value can lead to under fitting.

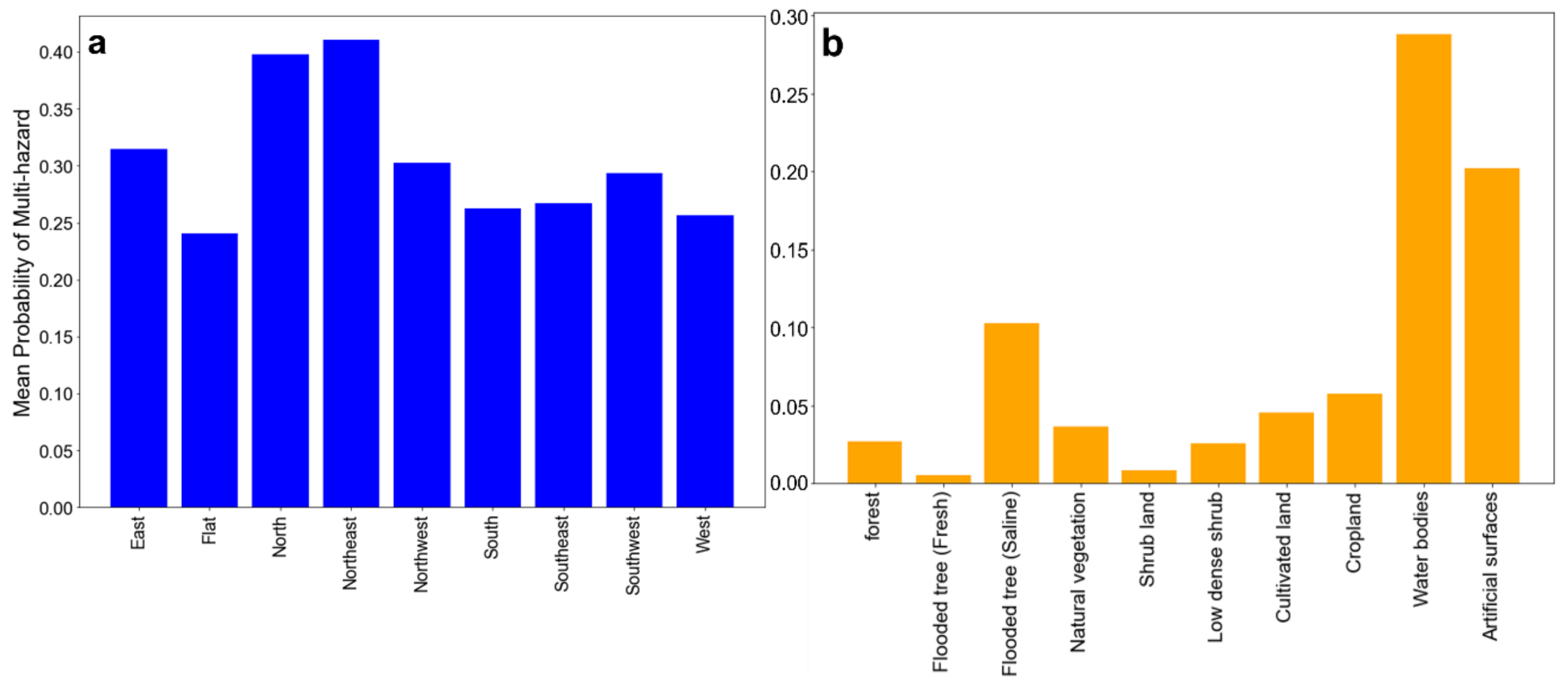
- **RF**
  - High predictive performance and ability to evaluate variable importance effectively.

RESULTS & DISCUSSION

- Comparison of NB, kNN, and RF predictive effectiveness.

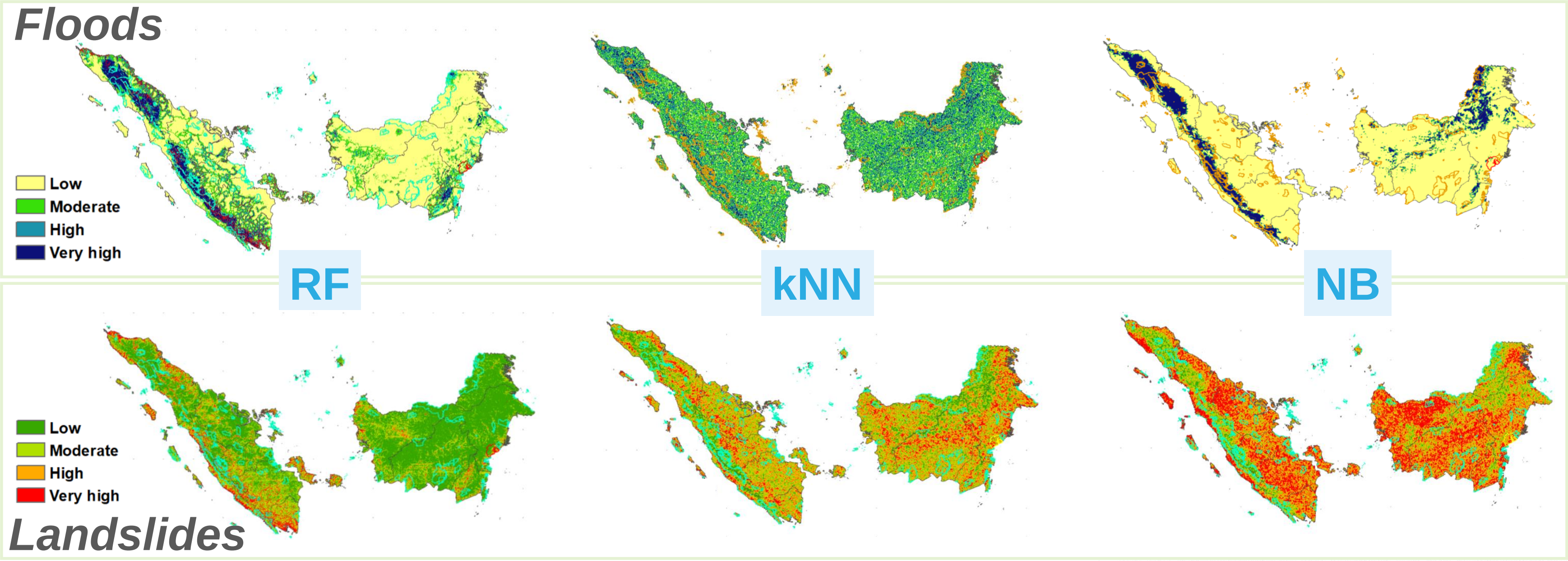
Hazard	Algorithms	AUC	RMSE	R^2
Floods	NB	0.415	0.717	-1.744
	kNN	0.827	0.405	0.126
	RF	**0.912**	**0.064**	**0.984**
Landsldies	NB	0.869	0.824	-2.620
	kNN	0.843	0.475	-0.204
	RF	**0.987**	**0.143**	**0.891**

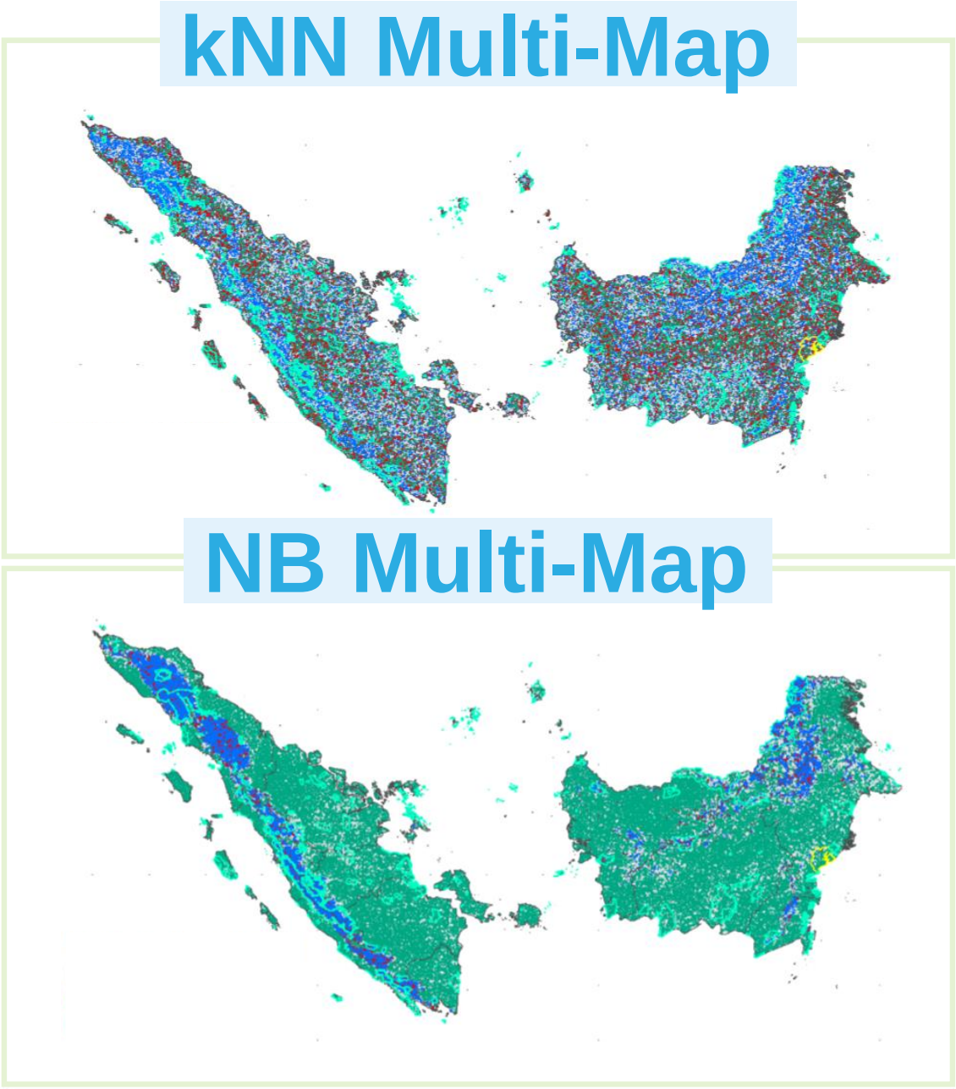
Note: RF model's accuracy against the kNN and NB models, revealing the RF model's superior reliability for disaster risk management.

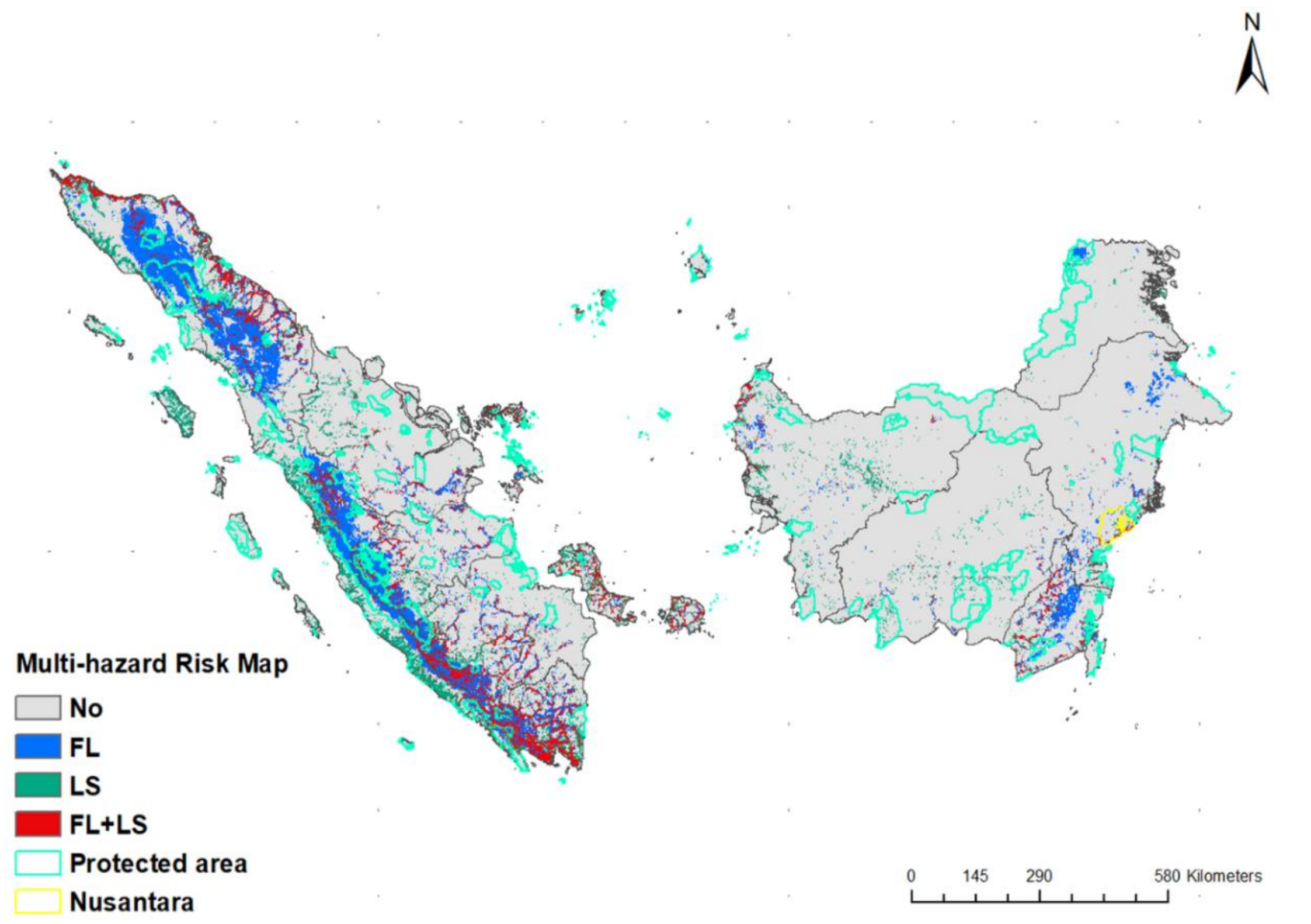


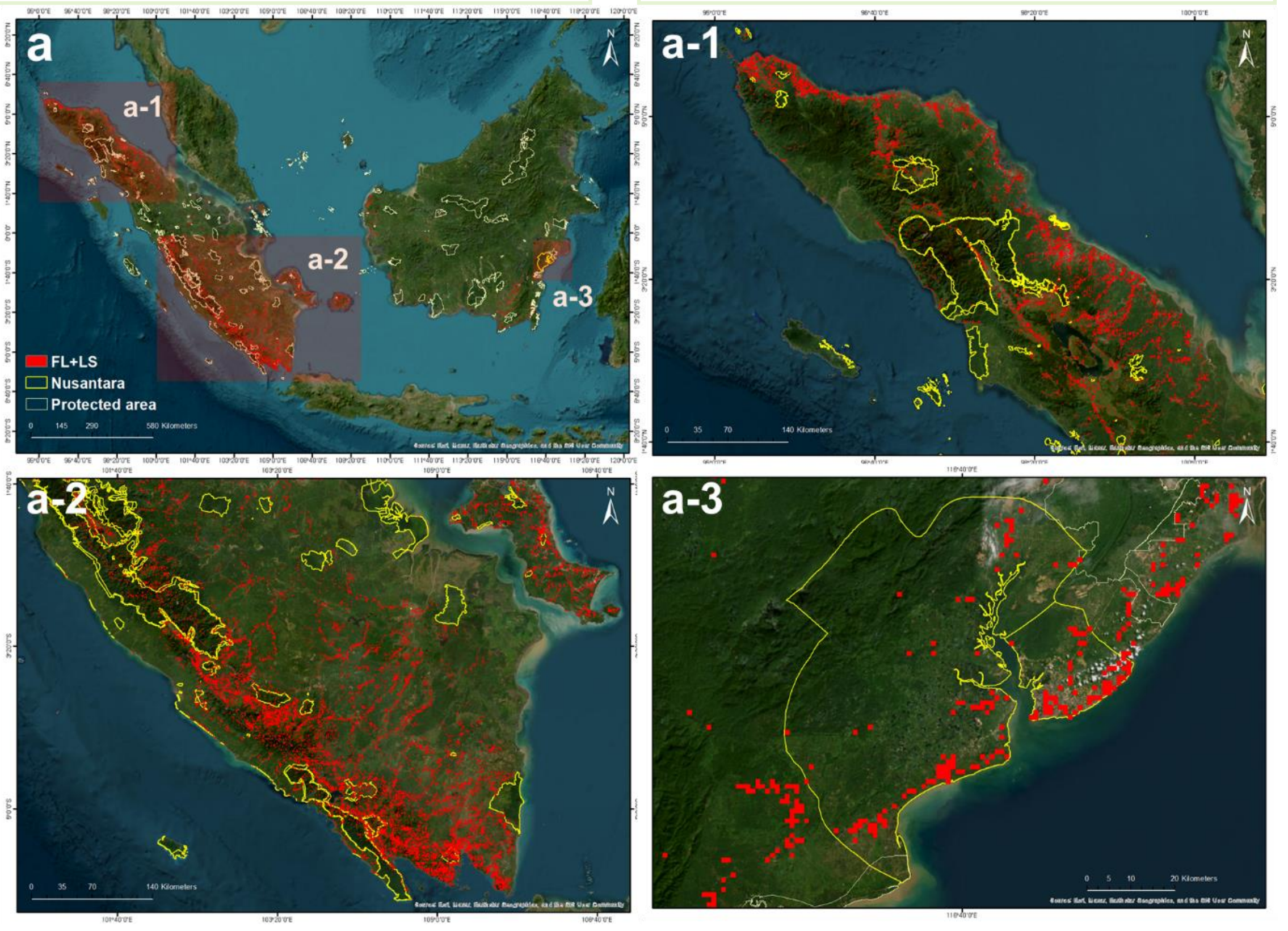
- Mean probability of multi-hazard by (a) slope aspect and (b) LULC type.

Note: Flood risk varies by land cover, with water bodies and urban areas having the highest risk, while forests and natural vegetation have low risk due to their hazard mitigation abilities. Land cover and topography significantly influence flood and landslide vulnerabilities.



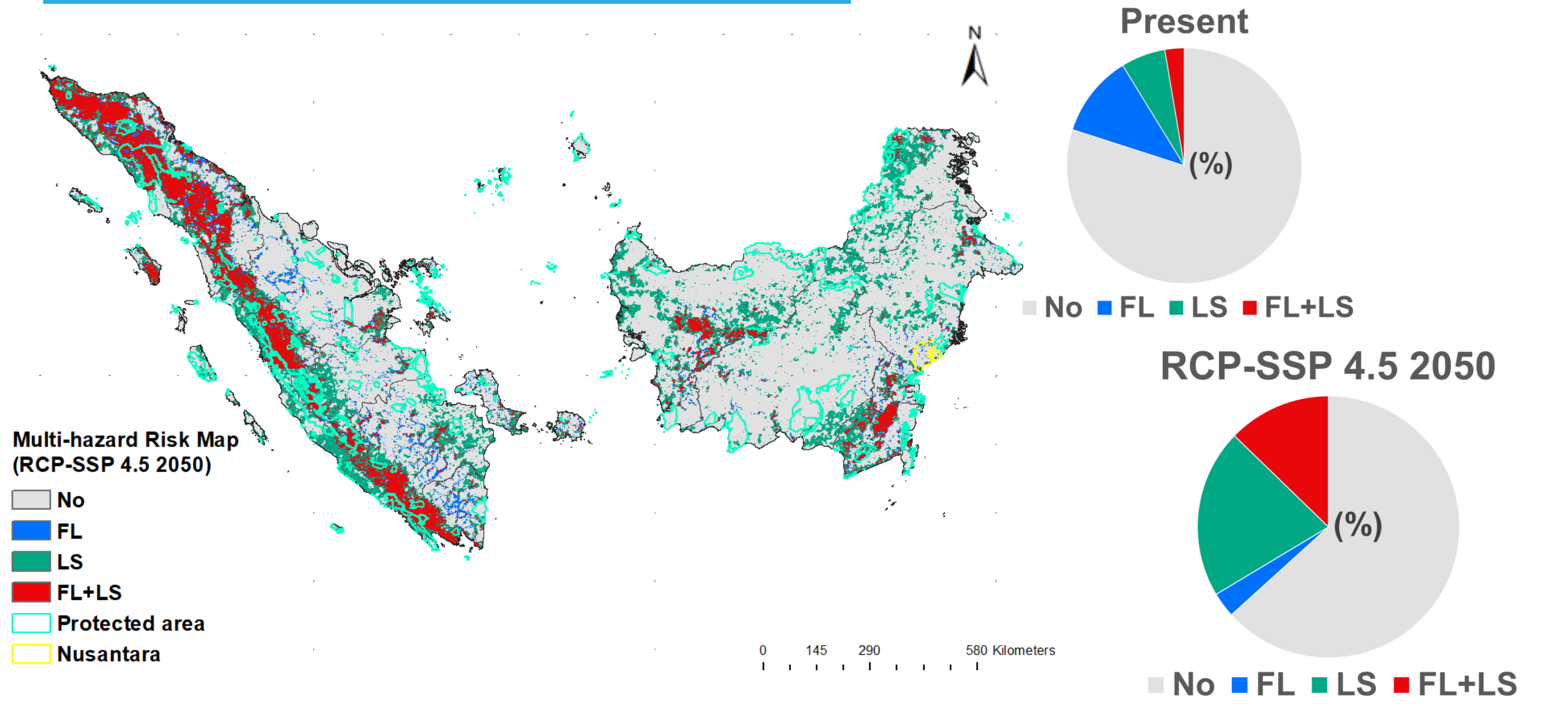






- Compared to the RF model, the kNN model overestimates single hazard risks, leading to potential unnecessary responses. The NB model overestimates landslide risks but underestimates multi-hazard risks, risking inadequate preparedness.
- Providing the Adaptive risk management strategies in multi-hazard areas, emphasizing the insufficiency of traditional single-hazard approaches and providing detailed hazard maps to guide effective resource allocation and long-term resilience planning.
- Emphasizing the cascading effects of hazards and the importance of using multi-hazard maps for infrastructure planning and resilience in disaster-prone regions.

CLIMATE SCENARIOS



❖Under the RCP-SSP 4.5 scenario for 2050; illustrating a shift in the proportions of different risk categories over time.

CONCLUSIONS

- Emphasizes the importance of understanding flood and landslide factors for sustainable development in Kalimantan and Sumatra, Indonesia.
- Using machine learning, it identified high-risk areas, with the RF model being the most effective.
- About 26.7% of the region is vulnerable to either hazard, and 16.8% to both. The new capital, Nusantara, is at higher multi-hazard risk.
- Integrated disaster management, considering land use and environmental factors, is crucial.
- Heavy rainfall and north-facing slopes increase risks, and targeted interventions in high-risk areas can enhance disaster preparedness and protect infrastructure and populations.

This work was supported by a grant from the Korea Environment Industry & Technology Institute (KEITI) through The Decision-Support System Development Project for Environmental Impact Assessment funded by the Korea Ministry of Environment (MOE) [grant number 2021003360002].