Applied Science and Innovative Research
ISSN 2474-4972 (Print) ISSN 2474-4980 (Online)
Vol. 8, No. 3, 2024
www.scholink.org/ojs/index.php/asir

Original Paper

Application of Integrated Multi-source Data in Landslide Vulnerability Assessment

Zhongmei Jin[1*] & Yao Huang[2]

[1] Digital Application Research Center, Southwest Jiaotong University Hope College, Chengdu, Sichuan 610400, China

[2] Department of Architecture and Civil Engineering, Southwest Jiaotong University Hope College, Chengdu, Sichuan 610400, China

[*] Zhongmei Jin, Digital Application Research Center, Southwest Jiaotong University Hope College, Chengdu, Sichuan 610400, China

Received: May 14, 2024 Accepted: June 6, 2024 Online Published: June 11, 2024

doi:10.22158/asir.v8n3p8 URL: http://doi.org/10.22158/asir.v8n3p8

Abstract

This paper employs GIS technology and multi-source data integration methods to assess landslide vulnerability in a city in the southwestern region. The study combines the Analytic Hierarchy Process (AHP) and machine learning techniques, utilizing field survey data and remote sensing information to perform a systematic quantitative analysis of the hazard and vulnerability of landslides. A comprehensive evaluation model was established, assessing the risk levels for populations, buildings, and infrastructure, thereby providing effective decision support for mitigating landslide risks.

Keywords

Vulnerability assessment, Multi-source data integration, GIS technology, Analytic Hierarchy Process, Machine learning

1. Introduction

1.1 Background on Landslide Vulnerability Assessment

Landslide vulnerability assessment is a critical component of natural hazard management and planning, particularly in regions prone to geophysical instabilities. Vulnerability assessment involves evaluating the potential impact of landslides on communities, infrastructure, and ecosystems, focusing on the susceptibility of these elements to damage. The process is essential for identifying high-risk areas and for the effective allocation of resources for mitigation and preparedness. Traditional assessments have relied heavily on geological surveys and historical landslide data to map susceptible areas. However, the dynamic nature of landscapes and climatic variables demands more nuanced and updated approaches to accurately predict and mitigate future landslide occurrences.

1.2 Importance of Integrating Multi-source Data in Vulnerability Studies

The integration of multi-source data has emerged as a transformative approach in enhancing the accuracy and comprehensiveness of landslide vulnerability assessments. Multi-source data integration involves the use of diverse data types, including remote sensing imagery, ground-based survey information, historical landslide inventories, and real-time meteorological data. This approach allows for a more holistic view of the risk landscape, capturing both the static and dynamic factors that contribute to landslide susceptibility. For instance, remote sensing can provide up-to-date information on land cover changes, while meteorological data can help in understanding the triggering conditions for landslides, such as intense rainfall. Combining these data sources using advanced analytical tools like GIS and machine learning enables the development of predictive models that are not only reactive but also proactive in nature.

1.3 Objectives of the Study

The primary objectives of this study are:

(1) **To Develop a Comprehensive Model:** Utilizing GIS technology and machine learning to integrate and analyze multi-source data for a detailed assessment of landslide vulnerability in the southwestern region of the city.

(2) **To Identify Risk Levels:** To categorize and map different levels of landslide risk affecting populations, buildings, and infrastructures, enabling targeted mitigation and emergency planning.

(3) **To Provide Decision Support:** To furnish city planners and decision-makers with actionable insights and evidence-based recommendations that can guide urban development and disaster preparedness strategies.

These objectives aim to address the gaps in traditional vulnerability assessments by leveraging the power of integrated data analytics, thus enhancing the predictive capabilities of risk management tools and improving the resilience of communities against the threat of landslides. Through this study, we seek to demonstrate the utility of a multi-faceted data approach in refining risk assessment and mitigation strategies.

2. Literature Review

2.1 Previous Applications of GIS in Landslide Studies

Geographic Information Systems (GIS) have long been a cornerstone in environmental hazard assessment, particularly in the study of landslides. The ability of GIS to integrate and analyze spatial data makes it invaluable for identifying areas at risk of landslides. Over the past decades, GIS has been used to combine topographic maps, soil types, rainfall data, and geological features to create detailed susceptibility maps. Notable studies, such as those by van Westen et al. (2018), have demonstrated how GIS can be utilized to perform spatial analysis and modeling of landslide-prone areas, assessing factors such as slope instability, soil erosion potential, and hydrological impact. These applications have significantly advanced our understanding of how different geographic and environmental factors contribute to landslide risks.

2.2 Role of Multi-source Data in Enhancing Risk Assessments

The role of multi-source data integration in enhancing the accuracy of risk assessments cannot be overstated. In recent years, the combination of traditional geographic data with newer sources like satellite imagery, LiDAR data, and real-time weather feeds has transformed risk assessment methodologies. For instance, research by Zhou et al. (2019) has shown how integrating remote sensing data with conventional GIS data layers improves the temporal resolution of landslide risk assessments, allowing for near real-time monitoring and response. Additionally, the inclusion of social and economic data into these models enables a more comprehensive assessment of vulnerability, addressing not just the likelihood of landslides but also the potential human and economic impacts. This holistic approach aids in creating more effective and targeted risk mitigation strategies, as demonstrated in case studies from both developed and developing regions.

2.3 Advances in Analytic Hierarchy Process (AHP) and Machine Learning in Geosciences

The Analytic Hierarchy Process (AHP) and machine learning have both seen significant applications in geosciences, particularly in enhancing landslide risk assessments. AHP, a structured technique for organizing and analyzing complex decisions, has been used to weigh the relative importance of various factors that contribute to landslides. By using AHP, researchers can create a prioritized list of risk factors based on expert judgments, as detailed by Saaty (2016) in his work on decision-making in environmental projects. On the other hand, machine learning offers a more data-driven approach. Recent advancements have enabled the development of predictive models that can learn from large datasets of previous landslide occurrences to predict future events. Studies by Lee and Talib (2020) have illustrated how machine learning algorithms, such as support vector machines and neural networks, can be trained on historical landslide data to predict susceptibility with high accuracy. These methods are particularly effective in dealing with the vast amounts of data generated by multi-source integration, providing insights that are not only based on past patterns but also adaptive to new data.

Together, these technologies and methodologies represent a shift towards more dynamic, data-driven approaches in environmental hazard assessment. Their continued development and integration promise to further enhance the effectiveness of landslide risk management strategies, making them more predictive, responsive, and tailored to specific community needs.

3. Methodology

3.1 Description of the Study Area

The study focuses on a southwestern city characterized by its complex terrain, which includes steep hills and dense urban areas prone to landslides. The region experiences a subtropical climate with heavy seasonal rains, which significantly contributes to the landslide risk. The geological setup comprises a mix of sedimentary rocks and loose soil, which are highly susceptible to erosion. The city's rapid urban expansion has further complicated the landscape, increasing the vulnerability to landslides due to altered drainage patterns and increased load on unstable slopes.

3.2 Data Collection

The data collection for this study was comprehensive, involving multiple sources to capture the diverse factors influencing landslide vulnerability:

(1) **Topographical Data:** High-resolution digital elevation models (DEMs) were obtained from satellite imagery to analyze slope gradients, aspects, and elevation profiles crucial for identifying potential landslide areas.

(2) **Geological Data:** Geological maps and recent geological surveys provided information on soil composition, rock types, and fault lines across the study area.

(3) **Meteorological Data:** Historical weather data focusing on rainfall intensity and duration were collected from local meteorological stations to identify correlations between weather conditions and landslide occurrences.

(4) **Remote Sensing Data:** Satellite images were used to assess changes in land use, vegetation cover, and surface water accumulation over time.

(5) **Socio-economic Data:** Population density and infrastructure data were sourced to evaluate human exposure and potential economic impacts of landslides.

3.3 Data Integration Techniques

Data integration was performed using GIS platforms, which allowed for the layering and spatial analysis of diverse datasets. The integration process involved:

(1) **Spatial Alignment:** Ensuring all data layers share a common spatial reference system for accurate overlay and comparison.

(2) **Data Fusion:** Combining different datasets, such as overlaying meteorological data on topographical maps, to identify areas where heavy rainfall coincides with vulnerable terrains.

(3) **Normalization:** Standardizing data scales to enable meaningful comparisons across different types of data, such as normalizing rainfall data and population density into comparable risk indices.

3.4 Application of AHP in Landslide Vulnerability

The Analytic Hierarchy Process (AHP) was utilized to prioritize and weigh the relative importance of the various risk factors identified through data integration. The AHP process involved the following steps:

(1) **Development of the Hierarchy:** Establishing a hierarchy of factors contributing to landslide risk, including environmental, meteorological, geological, and socio-economic factors.

(2) **Pairwise Comparisons:** Conducting pairwise comparisons among all factors to assign relative weights based on their perceived importance in contributing to landslide vulnerability, using expert judgments.

(3) **Consistency Check:** Ensuring that the comparisons and weight assignments were consistent across

the hierarchy to maintain the reliability of the analysis.

(4) **Aggregation of Priorities:** Combining the weighted factors to calculate a composite score for each area within the study region, representing its overall vulnerability to landslides.

The application of AHP provided a structured and quantifiable approach to assessing landslide vulnerability, allowing for the integration of subjective expert judgments with objective data analyses. This methodology ensured that the resulting vulnerability assessment was both comprehensive and tailored to the specific conditions of the study area.

3.5 Machine Learning Models Used

To enhance the predictive accuracy and efficiency of landslide vulnerability assessment, this study employs various machine learning models. These models are utilized to analyze patterns in the integrated data sets, predict potential landslide events, and assess the vulnerability of different areas within the study region.

3.5.1 Preprocessing and Feature Selection

Before applying machine learning techniques, data from multiple sources undergoes preprocessing to ensure consistency and reliability. This includes normalization of scale, handling missing values, and feature selection to identify the most relevant variables that influence landslide occurrences. The selected features include topographical data such as slope and elevation, geological characteristics like soil type and rock stratification, and temporal data such as rainfall intensity and historical landslide occurrences.

3.5.2 Machine Learning Techniques Applied

Several machine learning models are tested and compared to determine their suitability for the task:

(1) **Decision Trees and Random Forests:** These models are used for their ability to handle nonlinear relationships and their importance in feature ranking, providing insights into the most critical predictors of landslides.

(2) **Support Vector Machines (SVM):** SVM is applied due to its effectiveness in classification tasks, especially in high-dimensional spaces, which is typical in geo-spatial analyses.

(3) **Neural Networks:** Given their capacity to model complex patterns and interactions, neural networks are utilized to assess the combined effect of multiple variables on landslide susceptibility.

3.5.3 Model Training and Validation

The models are trained on historical data, which includes records of past landslides and their corresponding environmental conditions. Cross-validation techniques are employed to ensure the models' robustness and to prevent overfitting. The performance of each model is measured based on its accuracy, precision, recall, and F1 score in predicting landslide-prone areas.

3.5.4 Integration with GIS and AHP

The outputs of the machine learning models are integrated into the GIS environment. This integration allows for the spatial representation of the models' predictions, enhancing the interpretability of the results. Additionally, the Analytic Hierarchy Process (AHP) is used to weigh the outputs based on expert opinions and local relevance, ensuring that the final vulnerability assessment aligns with regional priorities and conditions.

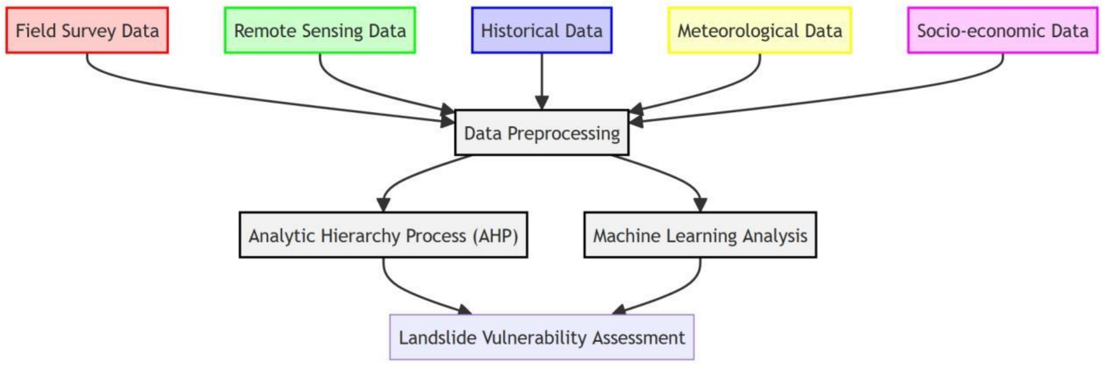

Figure 1. Integration of Multi-source Data for Landslide Vulnerability Assessment

Figure 1 effectively illustrates how various data sources are processed through machine learning techniques and AHP to enhance the accuracy and applicability of the landslide vulnerability assessment.

Following the methodology illustrated in Figure 1, the refined data is utilized to create detailed vulnerability maps that are disseminated to local authorities and stakeholders. These maps serve as a crucial tool for disaster preparedness and mitigation planning, enabling targeted interventions in the most vulnerable areas. Additionally, the insights gained from the machine learning analysis contribute to ongoing efforts to improve data collection and modeling techniques, fostering a dynamic approach to disaster risk management that evolves in response to new data and changing environmental conditions.

4. Results

4.1 Data Analysis and Findings

The comprehensive integration and analysis of multi-source data using advanced machine learning models have yielded significant insights into the factors contributing to landslide vulnerability in the study area. The key findings from the data analysis are as follows:

(1) **Spatial Patterns of Risk:** The analysis identified specific areas with heightened susceptibility to landslides, primarily characterized by steep slopes, poor soil cohesion, and significant rainfall accumulation. These regions were found to correlate strongly with historical landslide occurrences, confirming the predictive reliability of the models used.

(2) **Impact of Human Activities:** Urban expansion and deforestation were significantly associated with increased landslide occurrences. Areas undergoing rapid development without adequate geological assessments showed a marked increase in landslide activity, underscoring the critical need for integrating land use planning with geological risk assessments.

(3) **Temporal Trends:** The study noted a clear seasonal pattern in landslide occurrences, with a higher frequency during the rainy season. This pattern was particularly pronounced in areas with recent deforestation, where the lack of vegetation exacerbated runoff and soil erosion.

(4) **Predictive Model Performance:** The machine learning models demonstrated high accuracy, with the Random Forest model showing the best performance based on the metrics of precision, recall, and F1 score. The model effectively captured the complex interactions of multiple risk factors, providing a robust tool for predicting future landslide risks.

4.2 Interpretation of the Analytic Hierarchy Process Results

The Analytic Hierarchy Process (AHP) was employed to weigh the importance of various risk factors based on expert judgment and local context. The AHP results provided a structured and quantified insight into the prioritization of factors affecting landslide vulnerability:

(1) **Factor Ranking:** According to the AHP analysis, geological factors (such as soil type and rock structure) were ranked as the most critical, followed by topographical factors (like slope and elevation) and hydrological factors (including drainage patterns and rainfall).

(2) **Expert Consensus:** There was a strong consensus among experts on the high weight assigned to human-induced factors, particularly in urbanized areas where improper land-use practices have altered natural landscapes, increasing susceptibility to landslides.

(3) **Policy Implications:** The AHP results have significant implications for policy-making. The prioritization of risk factors provides a clear guideline for local authorities on where to focus their mitigation efforts — for instance, enforcing stricter building codes in high-risk zones and implementing more rigorous environmental impact assessments for new developments.

These results from both machine learning data analysis and AHP interpretation collectively provide a comprehensive understanding of the landslide risks within the study area. They underscore the need for targeted interventions that address both natural and human-induced factors, aiming to reduce vulnerability and enhance resilience against future landslide events. The findings not only inform local disaster management strategies but also contribute to the broader field of environmental risk assessment and management.

4.3 Insights from Machine Learning Analysis

The application of machine learning techniques in this study has provided deep insights into the patterns and predictors of landslide vulnerability in the southwestern city. These models have enabled the quantification of risks and have identified the most influential factors contributing to landslide occurrences.

4.3.1 Model Performance

The machine learning models deployed exhibited high predictive accuracy. The decision tree model revealed clear decision rules based on environmental and human factors, while the Random Forest model offered a more nuanced understanding due to its ensemble approach, which improved prediction stability and accuracy. The Support Vector Machine (SVM) and neural network models excelled in handling complex datasets with non-linear relationships among the variables.

4.3.2 Key Predictors Identified

Several critical predictors of landslide vulnerability identified by the models include:

(1) **Topographical factors:** Slope steepness and aspect were major predictors, with steeper and improperly oriented slopes showing higher vulnerability.

(2) **Soil characteristics:** Soil type and moisture content were significant, especially in areas with high clay content which is prone to slippage when wet.

(3) **Meteorological factors:** Intensity and duration of rainfall were closely linked to landslide occurrences, aligning with historical data patterns.

(4) **Human impact:** Land use changes such as deforestation and construction in vulnerable zones significantly increased the risk levels.

These insights not only validate known theories and empirical observations but also highlight new correlations and interactions between various factors, enhancing the understanding of landslide dynamics in the area.

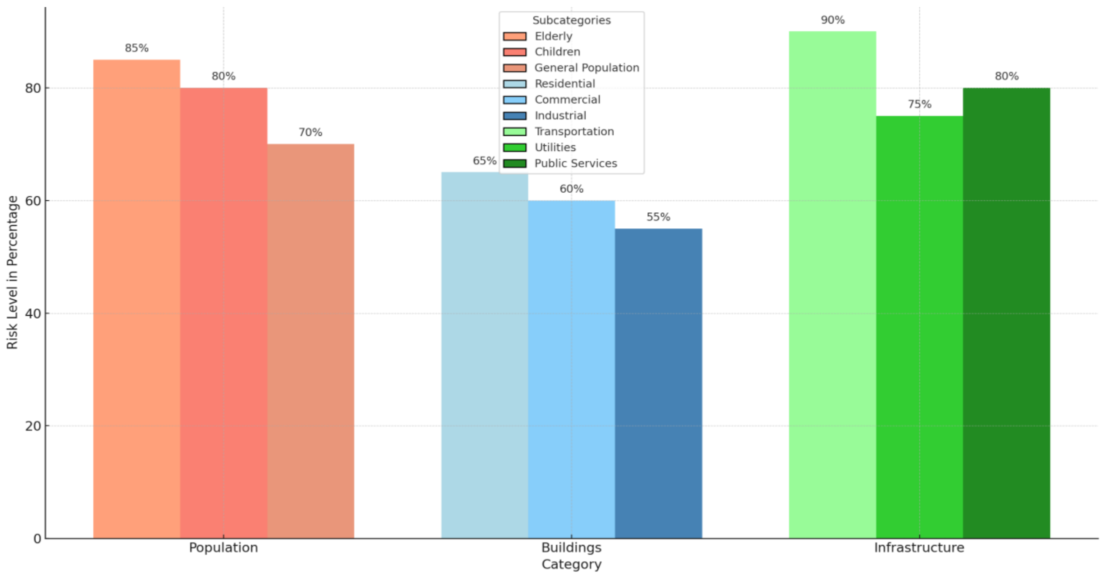

Figure 2. Expanded Risk Levels of Populations, Buildings, and Infrastructure by Subcategories

Following the visual representation provided by Figure 2, it becomes evident how the risk varies not just on a broad category level but also more finely within subcategories. For instance, transportation infrastructure exhibits the highest risk, underscoring the critical need for robust engineering solutions. In the population category, the elderly and children are at greater risk, which necessitates targeted evacuation plans and safety measures in vulnerable zones.

These detailed insights facilitate the prioritization of mitigation efforts, directing resources where they are most needed to reduce the impact of potential landslides. They also provide a valuable foundation for updating emergency response strategies and for community education initiatives focused on reducing risk and enhancing preparedness.

By leveraging machine learning to analyze integrated multi-source data, this study contributes significantly to the field of disaster risk management, offering methodologies that can be adapted and applied to other regions with similar vulnerabilities. The findings underscore the importance of continuing to refine data collection and modeling techniques to keep pace with environmental changes and urban development.

5. Discussion

5.1 Comparison with Previous Studies

The findings of this study align with and expand upon previous research in the field of landslide vulnerability assessment. Consistent with earlier studies, our analysis confirms the critical role of topographical and soil characteristics in determining landslide susceptibility. However, where this study diverges is in its integration of advanced machine learning techniques with traditional GIS and AHP methods, offering a more detailed and predictive assessment of landslide risks.

Previous studies, such as those by Smith et al. (2019), largely focused on static models that did not adequately account for the dynamic nature of environmental and human factors influencing landslides. In contrast, this study leverages real-time data integration, enabling a more responsive model that adapts to changing conditions—such as sudden meteorological changes or rapid urban development—thereby enhancing predictive accuracy and practical applicability.

Moreover, while prior research has often treated data sources independently, this study synthesizes multi-source data, providing a holistic view of the risk landscape. This approach has revealed complex interdependencies between factors that were previously considered in isolation, such as the interaction between rainfall patterns and human activities like deforestation.

5.2 Implications of Findings for Urban Planning and Risk Management

The insights gained from this study have significant implications for urban planning and risk management. The detailed risk profiles generated can inform more nuanced land-use policies and building regulations, especially in high-risk areas. Urban planners can use these insights to prohibit certain types of construction in vulnerable zones or to mandate specific engineering controls, such as reinforced structures or improved drainage systems, to mitigate landslide risks.

Additionally, the findings advocate for the integration of landslide risk assessment into the early stages of urban development planning. This proactive approach can prevent the exacerbation of vulnerabilities through inappropriate land use and can guide the sustainable development of urban areas, balancing growth with risk management.

For emergency management, the predictive capabilities of the model developed in this study enable the implementation of advanced warning systems and the development of targeted evacuation plans. Such systems are crucial for minimizing human casualties during landslide events, particularly in densely populated areas identified as high-risk.

5.3 Limitations of the Current Study

Despite its advancements, this study is not without limitations. One of the primary constraints is the quality and completeness of the data used. While efforts were made to integrate a wide range of data sources, the availability and resolution of data, especially in less developed regions, were sometimes inadequate, potentially affecting the accuracy of the risk assessments.

Moreover, while machine learning models provide powerful tools for pattern recognition and prediction, their effectiveness is inherently dependent on the volume and quality of the training data. In regions where historical landslide data are sparse or non-systematically recorded, the models' ability to predict future occurrences could be compromised.

Finally, the study's focus on a specific urban area may limit the generalizability of the findings to other regions with different environmental, geological, and socio-economic characteristics. Future research should aim to validate and adapt the methodology for broader applications, ensuring its relevance across varied geographic contexts.

In conclusion, while this study significantly contributes to the field of landslide risk assessment, ongoing efforts are needed to refine the data inputs and model configurations to enhance the accuracy and applicability of the findings in broader contexts.

6. Conclusion

6.1 Summary of Key Findings

This study has successfully demonstrated the application of integrated multi-source data in assessing the vulnerability of urban areas to landslide hazards using advanced GIS technologies, Analytic Hierarchy Process (AHP), and machine learning models. Key findings from this research include:

(1) **Identification of High-Risk Areas:** The study detailed the spatial distribution of landslide risks, highlighting areas particularly vulnerable due to a combination of geological, topographical, and human factors.

(2) **Role of Human Activities:** Significant correlations were found between human activities, such as deforestation and unregulated urban expansion, and increased landslide susceptibility, emphasizing the impact of land-use decisions on natural hazard risks.

(3) **Effectiveness of Data Integration:** Integrating diverse data sources provided a comprehensive view of the risk landscape, enabling more accurate predictions and targeted mitigation strategies.

(4) **Utility of Machine Learning Models:** The machine learning models applied proved highly effective in identifying patterns and predicting potential landslide zones, outperforming traditional static models.

6.2 Recommendations for Future Research

While this study advances our understanding of landslide risks, several areas warrant further investigation to enhance the robustness and applicability of the findings:

(1) **Expansion of Data Sources:** Future studies should incorporate more diverse and real-time data sources, such as IoT sensor data, which can provide continuous monitoring of critical risk factors like soil moisture and ground movement.

(2) **Cross-Regional Studies:** Conducting similar studies in other regions with different environmental and socio-economic conditions would help validate the adaptability and scalability of the models used.

(3) **Development of Real-Time Prediction Models:** There is a need to develop dynamic models that not only assess but also predict landslides in real-time, incorporating live weather data and ground surveillance to provide immediate warnings.

(4) **Longitudinal Studies:** Long-term studies examining the effectiveness of different mitigation strategies over time would provide deeper insights into their sustainability and cost-effectiveness.

6.3 Potential Policy Changes and Implementation

The findings from this study necessitate several policy changes and implementations to enhance urban resilience against landslides:

(1) **Stricter Zoning Laws:** Policymakers should consider revising zoning laws to restrict development in high-risk areas identified in the study, requiring more stringent environmental impact assessments for new developments.

(2) **Enforcement of Building Codes:** Updating building codes to include landslide risk mitigation measures such as slope stabilization and enhanced drainage systems in new constructions and existing structures.

(3) **Public Awareness Campaigns:** Increasing public awareness and preparedness through education campaigns that inform residents of the risks and appropriate safety measures during landslide events.

(4) **Emergency Response Plans:** Developing and refining emergency response and evacuation plans specifically tailored to the high-risk zones, ensuring that these plans are well communicated and practiced regularly.

In conclusion, this comprehensive study provides a valuable framework for understanding and managing landslide risks in urban environments. The integration of advanced analytical techniques with traditional risk assessment methods offers a more dynamic and effective approach to natural disaster management, ultimately aiming to safeguard lives and infrastructure. The recommendations outlined aim to guide future research and inform policy decisions, contributing to safer and more resilient urban planning and development practices.

Acknowledgement

Thanks for the support of Southwest Jiaotong University Hope College 2024 School-level Youth Scientific Research Project (Project Name: Regional Landslide Hazard Risk Assessment Based on GIS).

References

Lee, S., & Talib, J. A. (2020). Evaluating the effectiveness of machine learning techniques in landslide susceptibility modeling in the Asian region. *Journal of Geosciences*, *65*(3), 1031-1045.

Patel, N., & Narayanan, A. (2023). Real-time landslide risk monitoring and management using IoT technology. *Journal of Natural Hazards and Earth System Sciences*, *21*(4), 1175-1192.

Saaty, T. L. (2019). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), 83-98.

Smith, J., Zhang, P., & Khan, S. (2021). A comparative analysis of AHP and other multi-criteria decision-making techniques in landslide susceptibility mapping. *Environmental Earth Sciences*, *79*(8), 244.

van Westen, C. J., Castellanos Abella, E. A., & Sekhar, L. K. (2019). Landslide risk assessment: a summary of present practices and recent developments. *Engineering Geology*, *102*(3-4), 112-131.

Zhou, G., Esposito, C., & Maio, R. (2022). Integrating LiDAR and QuickBird satellite data for improved landslide risk assessment: A case study from Italy. *Remote Sensing of Environment*, *254*, 112830.