Article Type: Research

Landslide susceptibility analysis for Azershahr region using SVM and logistic regression methods

Elmira Ahadi[1], Daryoush Ahadi Rosta[2,*]

[1]Department of Civil Engineering and Geotechnics, Islamic Azad University, Aras Branch, Hadishahr 5441734335, Iran
[2]Department of Civil Engineering, Aeen Kamal Higher Education Institute, Urmia 5715938911, Iran
*Corresponding author: ahadi.daryoush.85@gmail.com

Received: 28 May 2024 / Accepted: 28 July 2024 / Published: 18 December 2024

***Abstract*:** Landslides are one of the most destructive natural hazards, posing significant threats to infrastructure, human life, and economic activities, particularly in mountainous regions. This study focuses on assessing landslide susceptibility in Azarshahr County, East Azerbaijan Province, Iran, using a combination of Support Vector Machine (SVM) and Logistic Regression (LR) techniques integrated with Geographic Information Systems (GIS). A total of 12 factors influencing landslide occurrence, including slope aspect, distance to the cities, elevation, drainage density, faults' unsafe radius, rainfall, distance to the river, distance to roads, slope angle, temperature, and weathering, were analyzed. The study employed advanced GIS-based spatial analysis and machine learning methodologies to classify landslide-prone areas into distinct susceptibility zones ranging from very low to very high risk. Results demonstrated the superior predictive capability of the SVM model for handling nonlinear relationships, while LR provided insights into the relative contributions of each factor. The susceptibility map revealed that significant portions of Azarshahr County are categorized as moderate to high-risk zones, particularly in areas with steep slopes and high rainfall. The outcomes offer valuable insights for land-use planning, disaster mitigation strategies, and community safety initiatives. This study highlights the effectiveness of integrating machine learning techniques with GIS for natural hazard assessment, providing a replicable framework for landslide susceptibility mapping in other vulnerable regions.

Keywords: Landslide susceptibility, Geohazards, Logistic regression, Machine learning, Azershahr.

I. INTRODUCTION

Landslides are among the most devastating natural disasters, causing significant destruction to infrastructure, agriculture, and human settlements worldwide. They occur as a result of the gravitational movement of rock, soil, or debris down a slope, often triggered by natural factors such as heavy rainfall, earthquakes, or volcanic activity. Human-induced activities like deforestation, mining, and construction also play a significant role in exacerbating landslide risks (Marjanović et al., 2011). Understanding landslide susceptibility is critical not only for minimizing the loss of life and property but also for ensuring sustainable development in vulnerable regions (Nikoobakht et al., 2022).

Generally, landslide susceptibility refers to the likelihood of landslide occurrence in a specific area based on inherent and external conditions (Huang & Zhao, 2018). It is a crucial component of disaster risk reduction strategies, particularly in regions with high population densities or valuable infrastructure (Ermini et al., 2005). As landslides can occur without warning, susceptibility studies provide a proactive approach by identifying potential high-risk areas and allowing for preemptive measures. Such studies are integral for policymakers, urban planners, and environmental managers aiming to safeguard communities and optimize land use (Zêzere, 2002).

In recent decades, Geographic Information Systems (GIS) and remote sensing technologies have emerged as indispensable tools in landslide susceptibility mapping. These technologies allow for the collection, integration, and analysis of diverse spatial data, enabling researchers to explore the complex interactions between environmental, geological, and climatic factors (Süzen & Doyuran, 2004). With GIS, vast datasets can be visualized and processed efficiently, while remote sensing provides high-resolution imagery to capture topographical and land-use changes. Together, these tools enable a systematic approach to identifying landslide-prone zones with unprecedented accuracy (Azarafza et al., 2018). As known, landslide occurrences are influenced by a multitude of factors that vary depending on the geographic and climatic context. Common contributing factors include slope gradient, soil type, rock structure, land cover, precipitation levels, and seismic activity (Erener & Düzgün, 2012). For instance, steep slopes and weak soil cohesion significantly increase the likelihood of landslides during intense rainfall. Additionally, human activities such as deforestation and improper land-use practices further exacerbate the risk by altering natural drainage patterns and destabilizing slopes. The interplay of these factors is critical for understanding and predicting landslide events (Reichenbach et al., 2018). Traditional statistical methods have long been used for landslide susceptibility mapping; however, their limitations in capturing complex relationships between variables have paved the way for machine learning techniques. Machine learning algorithms, such as Support Vector Machines (SVM) and Logistic Regression (LR), have proven effective in analyzing nonlinear and multidimensional data (Huang & Zhao, 2018). SVM, in particular, excels in classifying landslide-prone areas by identifying subtle patterns in the data. LR, on the other hand, offers a simpler and more interpretable approach, making it suitable for binary classifications like determining landslide presence or absence (Pourghasemi et al., 2018).

The use of SVM and LR in landslide studies offers complementary strengths. SVM is adept at handling large and complex datasets with nonlinear relationships, making it highly effective for regions with diverse geological and climatic conditions. LR, while simpler, provides clear insights into the relative importance of various contributing factors. By combining these methods, researchers can achieve both high accuracy and a deeper understanding of the factors influencing landslide susceptibility. This hybrid approach also allows for the validation and comparison of results, ensuring robustness in the findings (Kalantar et al., 2018). Despite advancements in technology, landslide susceptibility mapping is not without challenges. Data availability and quality often pose significant barriers, particularly in remote or underdeveloped regions. The heterogeneity of geological formations and climatic conditions also complicates the modeling process, requiring careful calibration and validation (Pourghasemi et al., 2018). Furthermore, the dynamic nature of landslides, influenced by changing environmental and anthropogenic factors, necessitates continuous monitoring and updating of susceptibility maps to remain relevant (Süzen & Doyuran, 2004).

This study focuses on assessing landslide susceptibility in Azarshahr County by integrating SVM and Logistic Regression models within a GIS framework. Azarshahr County, situated in the East Azerbaijan Province of Iran, provides an ideal case study for landslide susceptibility analysis due to its diverse topography and climatic conditions (see Figure 1). The region is characterized by steep slopes, loose soil formations, and periodic heavy rainfall, all of which contribute to a high incidence of landslides. Additionally, the area's expanding agricultural and urban activities further increase vulnerability by disrupting natural stability. Understanding the susceptibility patterns in Azarshahr is crucial for the safety and development of local communities. The research aims to identify the most influential factors contributing to landslide occurrence and map high-risk zones with precision. By utilizing GIS for spatial analysis and remote sensing for data acquisition, the study ensures a comprehensive and systematic approach. Validation of the results will be conducted through field surveys and statistical evaluation, ensuring reliability and accuracy in the findings. The outcomes of this research have significant implications not only for Azarshahr County but also for similar regions worldwide. The integration of advanced modeling techniques with GIS provides a replicable framework for landslide susceptibility studies in other vulnerable areas. Additionally, the findings can guide local authorities in implementing targeted mitigation measures, such as slope stabilization, reforestation, and land-use planning. Ultimately, this study contributes to the broader goals of disaster risk reduction and sustainable development in landslide-prone regions.

II. Materials and Method

Support Vector Machines (SVM) and Logistic Regression (LR) are widely used in landslide susceptibility assessments due to their ability to handle complex data patterns and provide interpretable results (Süzen & Doyuran, 2004). SVM works by finding the optimal hyperplane that separates landslide-prone and non-prone areas based on input factors such as slope angle, lithology, and rainfall (Pourghasemi et al., 2018). It uses key data points, called support vectors, to define this boundary, ensuring robustness and precision. For non-linear relationships often observed in landslide data, SVM employs kernel functions to project data into higher dimensions, enabling the identification of intricate patterns. This makes SVM highly effective for mapping susceptibility in diverse terrains, where geological and environmental factors interact in complex ways (Huang & Zhao, 2018). LR complements landslide studies by estimating the probability of landslide occurrence based on linear relationships between predictors and outcomes. It uses a logistic function to map input variables like soil type, land use, and drainage density; to a probability scale, offering a straightforward classification into landslide-prone or safe zones. Logistic Regression's simplicity and interpretability make it a reliable choice for preliminary assessments or as a benchmark for comparing advanced models.

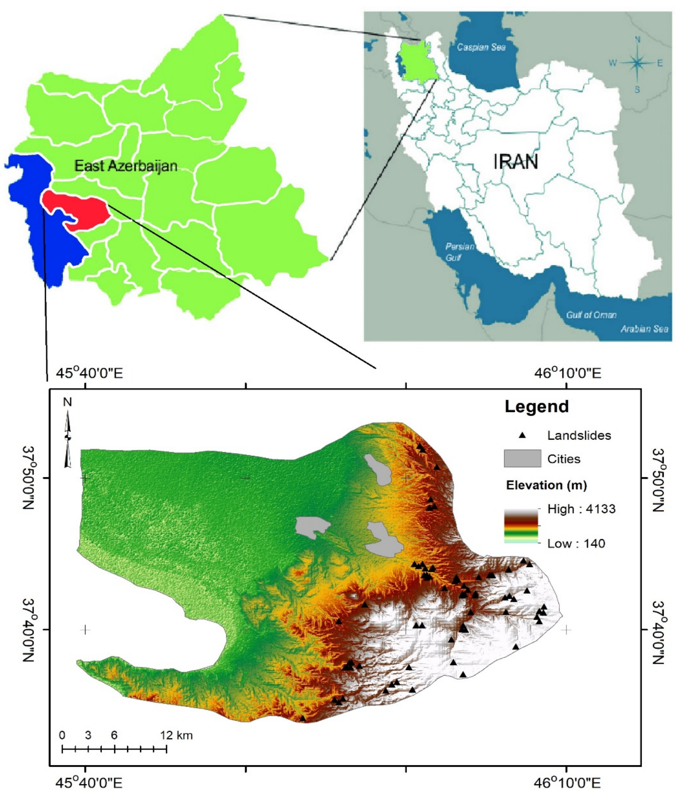

Fig. 1 Location of studied area

https://doi.org/10.22034/CGEL.1.2.e100022

The study focuses on Azarshahr, a region with diverse geological and environmental conditions, making it prone to landslides. The first step involves collecting spatial and environmental datasets critical for assessing landslide susceptibility. A total of 12 factors influencing landslide occurrence, including slope aspect, distance to the cities, elevation, drainage density, faults' unsafe radius, rainfall, distance to the river, distance to roads, slope angle, temperature, and weathering (Figure 2). Data is obtained from remote sensing, geological maps, and field surveys. A landslide inventory is compiled from historical records and on-site investigations, which is crucial for model training and validation. The inventory is divided into two sets: 70% for training the models and 30% for testing. Before modeling, the collected data undergoes preprocessing, such as normalization to ensure consistency in scale and statistical checks like variance inflation factor (VIF) analysis to address multicollinearity among factors. Table 1 is provided the triggering factors impact and inflation factor assigned for each of the factors accordingly (Figure 3).

The SVM model is designed to classify areas as landslide-prone or non-prone based on conditioning factors. Given the complex and non-linear nature of landslide-triggering processes, an RBF (radial basis function) kernel is chosen. Parameters such as the penalty term (C) and kernel coefficient (γ) are optimized through grid search and cross-validation to ensure the model's accuracy. The training phase involves the algorithm identifying a hyperplane that best separates the two classes while maximizing the margin. Once trained, the SVM model is applied to the entire study area to predict landslide susceptibility. The results are integrated with GIS to produce a susceptibility map, highlighting high, moderate, and low-risk zones across Azarshahr. LR is implemented as a complementary method to SVM, providing a probabilistic approach to landslide prediction. The model establishes a logit relationship between landslide occurrence and conditioning factors, estimating the probability of landslide events.

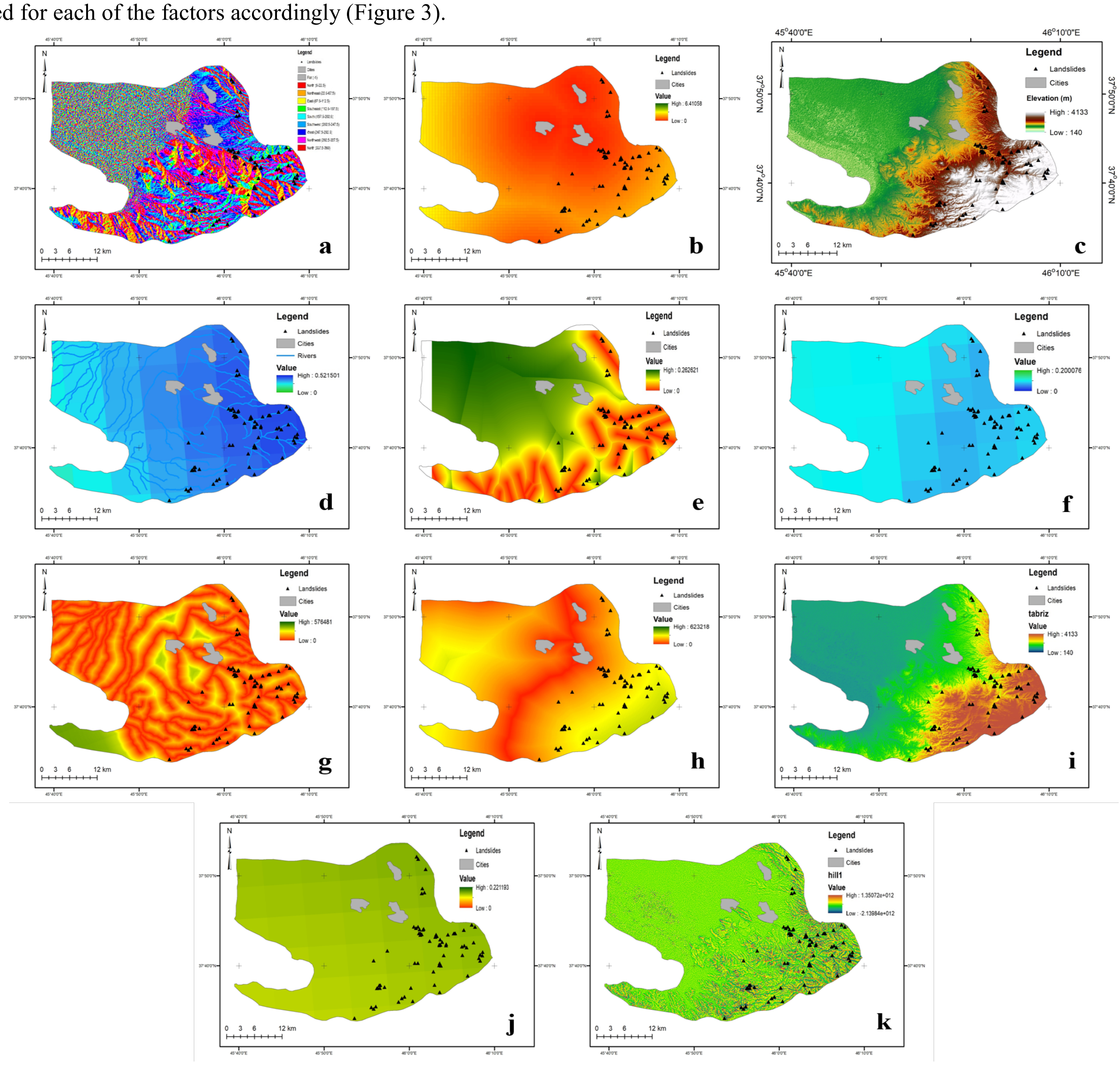

Fig. 2 The triggering factor maps for landslides in studied region: (a) slope aspect, (b) distance to the cities, (c) elevation, (d) drainage density, (e) faults' unsafe radius, (f) rainfall, distance to the river, (g) distance to roads, (h) slope angle, (i) curvature, (j) temperature (k) weathering

https://doi.org/10.22034/CGEL.1.2.e100022

Table 1 The effective triggering factors with impact on susceptibility analysis

Class	Triggering factors	Resolution	Data source	VIF index
Morphologic	Slope aspect	± 30 m	DEM	1.39
	Elevation	± 30 m	DEM	1.17
	Slope angle	± 30 m	DEM	0.92
	Curvature	± 30 m	DEM	0.85
Geologic	Drainage density	± 30 m	Geological data	1.35
	Faults' unsafe radius	± 30 m	Landsat TM, IWRM*	1.12
	Distance to the river	± 30 m	Landsat TM, IWRM*	1.44
	Weathering	± 30 m	Geological data	1.52
Climatologic	Rainfall	± 30 m	Landsat TM, IMO†	1.63
	Temperature	± 30 m	Landsat TM, IMO†	1.09
Human-activity	Distance to roads	± 30 m	DEM, Google Map	1.85
	Distance to the cities	± 30 m	DEM, Google Map	2.00

Note: *Iran Water Resources Management Company (IWRM); †Iran Meteorological Organization (IMO)

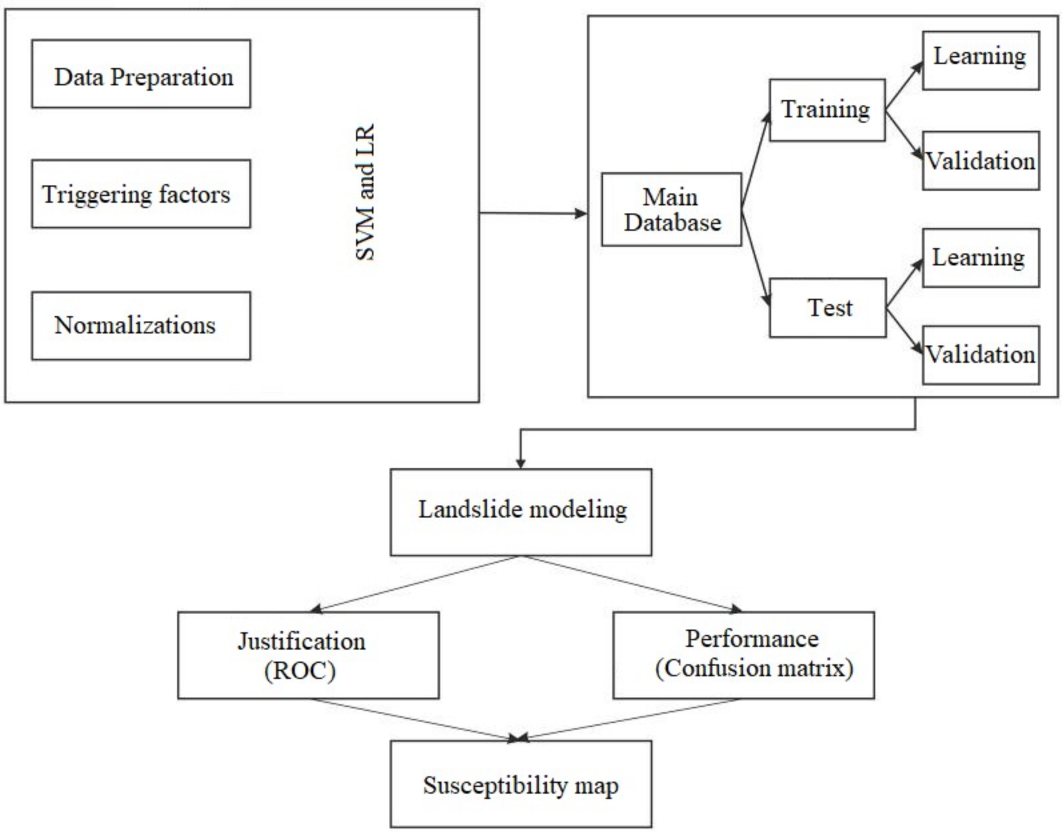

Fig. 3 The process flowchart for the studied methodology

The regression coefficients for each variable are determined using maximum likelihood estimation (MLE), ensuring a statistically robust model. The probabilities calculated for each spatial unit are classified into susceptibility classes based on pre-defined thresholds (Pourghasemi et al., 2018). The susceptibility map generated using LR provides a straightforward and interpretable overview of landslide risks, serving as a benchmark for model comparisons and practical applications. The effectiveness of both SVM and LR models is evaluated using accuracy metrics, including the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC). These metrics indicate the models' ability to distinguish between landslide-prone and non-prone areas. Validation against the test dataset and field observations ensures the reliability of the predictions.

The ROC curve is a powerful tool for evaluating the performance of predictive models in landslide susceptibility analysis (Frattini et al., 2010). It visually represents the trade-off between a model's true positive rate (sensitivity) and false positive rate (1-specificity) across different classification thresholds. By plotting these rates, the ROC curve helps assess how well a model distinguishes between landslide-prone and non-prone areas. A model that performs perfectly will have a curve that closely hugs the top-left corner of the plot, indicating high sensitivity with minimal false positives. In landslide susceptibility studies, the AUC derived from the ROC curve is a key metric for quantifying model accuracy (Cantarino et al., 2019). AUC values range from 0 to 1, with higher values indicating better discrimination. For example, an AUC of 0.85 means the model has an 85% chance of correctly distinguishing

https://doi.org/10.22034/CGEL.1.2.e100022

between landslide-prone and non-prone areas (Cantarino et al., 2019). This makes the ROC-AUC particularly useful for comparing different models, such as SVM and LR, in terms of their predictive performance (Das et al., 2010). In practice, the ROC curve helps researchers and planners identify the most reliable model for creating accurate susceptibility maps, ensuring effective risk assessment and mitigation strategies.

The comparative analysis helps identify the strengths of each model. For instance, SVM may excel in handling non-linear relationships, while LR provides simplicity and transparency, aiding decision-makers in understanding the factors contributing to landslide susceptibility. The susceptibility maps derived from SVM and LR models are essential tools for disaster risk management and urban planning in Azarshahr. High-risk zones identified by the models can guide the prioritization of mitigation strategies, such as slope stabilization, improved drainage systems, and reforestation efforts. Urban planners can use these maps to avoid high-risk areas during infrastructure development and allocate resources effectively. The integration of advanced machine learning models like SVM with traditional approaches such as LR ensures a comprehensive understanding of landslide risks, promoting sustainable land-use planning and reducing potential damages in the region.

SVM and LR are highly effective tools for landslide susceptibility analysis in Azarshahr due to their ability to handle diverse datasets and deliver reliable predictions. Azarshahr's landslide-prone terrain involves complex interactions between various factors such as slope, lithology, rainfall, and proximity to faults and rivers. SVM excels in capturing non-linear relationships within these datasets by employing kernel functions, allowing it to model the intricate patterns associated with landslide triggers. Its capacity to maximize the margin between landslide-prone and non-prone areas ensures precise classification, making it ideal for the region's diverse geological and environmental conditions. LR, on the other hand, offers a probabilistic approach that is straightforward and interpretable, making it particularly valuable for understanding the contribution of each factor to landslide occurrences. In Azarshahr, where decision-makers need clear and actionable insights, LR's ability to quantify the influence of variables like slope angle, drainage density, and rainfall on landslide probability is crucial. It simplifies the decision-making process by providing a clear threshold for susceptibility classification, enabling stakeholders to prioritize mitigation strategies effectively. Combining SVM and LR in landslide analysis for Azarshahr ensures a balanced approach. While SVM provides high accuracy and handles complex data structures, LR adds value through its simplicity and transparency. Together, they create robust susceptibility maps that not only highlight high-risk zones but also offer insights into the underlying causes of landslides. This dual-method approach enhances disaster preparedness and supports sustainable urban planning, helping to mitigate landslide risks in the region effectively.

III. Results and Discussion

As previously mentioned, this study aimed to produce a landslide susceptibility map for Azarshahr County, located in East Azerbaijan Province in the northwest of Iran. The region is characterized by complex geological formations, diverse topography, and varying climatic conditions, making it prone to landslides. By analyzing the interplay of environmental and anthropogenic factors, this research seeks to identify high-risk zones and provide critical insights for disaster mitigation and land-use planning. Azarshahr County, located in East Azerbaijan Province in northwest Iran, is a region highly susceptible to landslides due to its unique geological, topographical, and climatic characteristics. The county is situated in a mountainous area with steep slopes, diverse lithological formations, and active tectonic zones. These factors, combined with frequent rainfall and temperature variations, create a favorable environment for slope instability. Human activities, such as road construction, deforestation, and improper land use, further exacerbate the risk of landslides in this region. Historically, Azarshahr has experienced at least 12 significant landslides, which have caused considerable damage to infrastructure, agricultural land, and local communities. These events highlight the region's vulnerability and the need for effective landslide susceptibility assessments. By understanding the factors that contributed to these past incidents, researchers can better predict and mitigate future landslides, ensuring the safety of residents and the sustainability of the local environment.

Landslide susceptibility mapping is crucial for Azarshahr County due to its high vulnerability to landslides, which pose significant threats to infrastructure, agriculture, and human safety. The region's steep terrain, active fault lines, and variable climatic conditions make it essential to identify high-risk zones for effective disaster risk management and sustainable land use planning. To achieve this, we employed SVM and LR, two powerful predictive modeling techniques, to analyze the relationships between environmental and anthropogenic factors and landslide occurrences. These models allowed us to create accurate susceptibility maps, providing vital insights for local authorities and planners to prioritize mitigation strategies and minimize potential damages. The methodology of model implementation has been discussed previously and process flowchart has been provided in Figure 3.

Figure 4 presents the landslide susceptibility maps for the studied region, developed using SVM and LR. These maps highlight the spatial distribution of landslide risks across Azarshahr County, providing a clear visual representation of high and low-risk zones. The maps are instrumental in understanding the county's varying susceptibility levels, offering critical insights for disaster management and urban planning. The eastern part of Azarshahr County is classified as having high to very high landslide susceptibility. This area aligns with locations where 12 historical landslides have occurred, emphasizing its vulnerability. The combination of steep slopes, proximity to active fault lines, and high rainfall in this region makes it prone to landslides. The overlap between high-risk zones and past events also validates the accuracy of the models used, showcasing their effectiveness in predicting landslide-prone areas.

In contrast, the western part of the county, located within the Urmia Lake plain, is classified as having low to very low landslide susceptibility. This classification is attributed to the relatively flat terrain, stable geological formations, and lower influence of landslide-triggering factors in this area.

https://doi.org/10.22034/CGEL.1.2.e100022

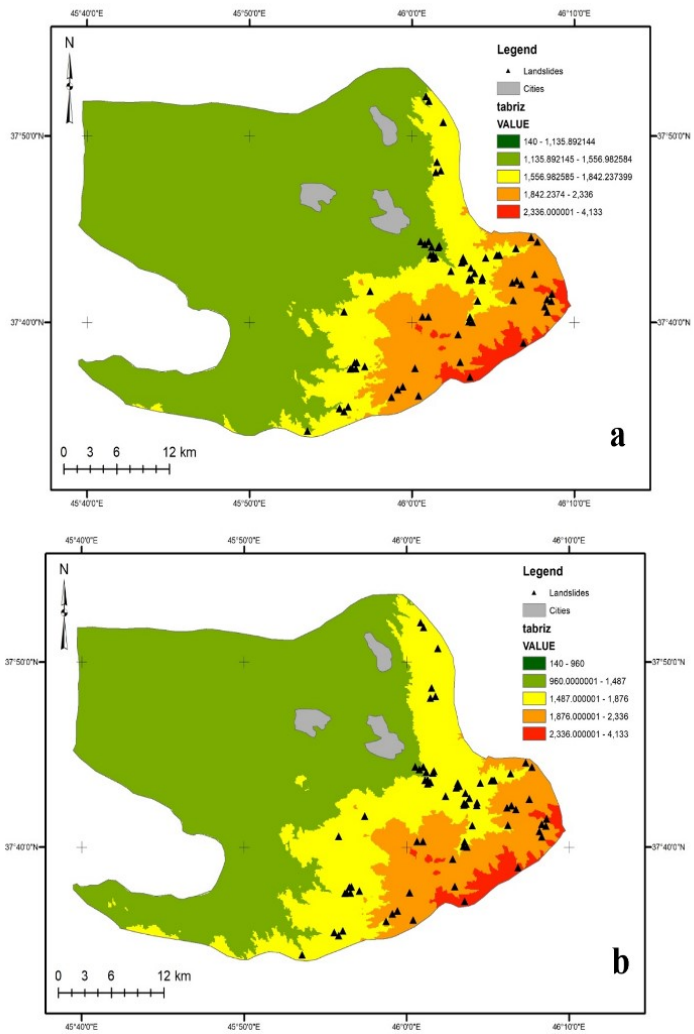

Fig. 4 Landslide susceptibility maps for the study area were developed using two models: (a) SVM, (b) LR

The minimal risk in this region underscores the importance of terrain and environmental stability in reducing landslide likelihood. The differentiation between high-risk zones in the east and low-risk areas in the west demonstrates the models' ability to capture the spatial variability of landslide susceptibility. These maps not only highlight regions requiring urgent mitigation measures but also provide a foundation for sustainable land use planning. By focusing on high-risk zones and minimizing development in these areas, local authorities can reduce the impact of future landslides and enhance the region's overall resilience.

IV. Conclusion

This study successfully developed landslide susceptibility maps for Azarshahr County using Support Vector Machine (SVM) and Logistic Regression (LR), providing valuable insights into the spatial distribution of landslide risks in the region. The results highlight the importance of predictive modeling in identifying areas at high risk of landslides and the need for effective mitigation measures. These maps serve as essential tools for disaster risk reduction and informed decision-making in land-use planning and urban development. The eastern part of Azarshahr County was identified as having high to very high landslide susceptibility, corroborated by the occurrence of 12 historical landslides in the region. This area's vulnerability is influenced by its steep slopes, active fault lines, and higher precipitation levels, which act as primary triggering factors for landslides. The models' accurate prediction of these zones demonstrates their reliability and potential for application in similar high-risk areas worldwide. In contrast, the western region, located in the flat Urmia Lake plain, exhibited low to very low susceptibility to landslides. The geological stability and minimal slope angles in this area significantly reduce the likelihood of landslide events. This contrast between the east and west highlights the influence of topography, geology, and hydrology on landslide susceptibility, underscoring the importance of considering these factors in regional risk assessments. Overall, the combination of SVM and LR provided a comprehensive approach to landslide susceptibility mapping, balancing advanced computational techniques with interpretability. The findings underscore the need for targeted mitigation strategies in high-risk zones while supporting sustainable development in safer areas. By implementing the insights gained from this study, local authorities can minimize future landslide risks, safeguard communities, and promote long-term resilience in Azarshahr County.

Acknowledgment

We would like to express our sincere gratitude to the editorial team for their valuable feedback and guidance throughout the review process. Additionally, we extend our thanks to the reviewers for their meticulous attention to detail and constructive suggestions that greatly improved the quality of this manuscript. Your contributions have been instrumental in shaping this work.

Authors' contributions

Elmira Ahadi and Daryoush Ahadi Rosta conducted the main data analysis, contributed to the data collection, preprocessing, and interpretation, and were responsible for drafting the initial manuscript. Daryoush Ahadi also, assisted in the development of the methodology and performed validation checks, provided supervision, conceptual guidance, and critical revision of the manuscript. All authors read and approved the final manuscript.

Conflict of interest

The authors have not disclosed any competing interests.

REFERENCES

Azarafza M., Ghazifard A., Akgün H., Asghari-Kaljahi, E. (2018). Landslide susceptibility assessment of South Pars Special Zone, southwest Iran. *Environmental Earth Sciences*, 77, 1-29. https://doi.org/10.1007/s12665-018-7978-1.

Cantarino I., Carrion M.A., Goerlich F., Ibañez, V.M. (2019). A ROC analysis-based classification method for landslide susceptibility maps. *Landslides*, 16, 265-282. https://doi.org/10.1007/s10346-018-1063-4.

Das I., Sahoo S., van Westen C., Stein A., Hack R. (2010). Landslide susceptibility assessment using logistic regression and its comparison with a rock mass classification system, along a road section in the northern Himalayas (India). *Geomorphology*, 114(4), 627-637. https://doi.org/10.1016/j.geomorph.2009.09.023.

Erener A., Düzgün H. (2012). Landslide susceptibility assessment: what are the effects of mapping unit and mapping method?. *Environmental Earth Sciences*, 66, 859-877. https://doi.org/10.1007/s12665-011-1297-0.

Ermini L., Catani F., Casagli N. (2005). Artificial neural networks applied to landslide susceptibility assessment. *Geomorphology*, 66(1-4), 327-343. https://doi.org/10.1016/j.geomorph.2004.09.025.

Frattini P., Crosta G., Carrara A. (2010). Techniques for evaluating the performance of landslide susceptibility models. *Engineering Geology*, 111(1-4), 62-72. https://doi.org/10.1016/j.enggeo.2009.12.004.

Huang Y., Zhao L. (2018). Review on landslide susceptibility mapping using support vector machines. *Catena*, 165, 520-529. https://doi.org/10.1016/j.catena.2018.03.003.

Kalantar B., Pradhan B., Naghibi S.A., Motevalli A., Mansor S. (2018). Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). *Geomatics, Natural Hazards and Risk*, 9(1), 49-69. https://doi.org/10.1080/19475705.2017.1407368.

Marjanović M., Kovačević M., Bajat B., Voženílek V. (2011). Landslide susceptibility assessment using SVM machine learning algorithm. *Engineering Geology*, 123(3), 225-234. https://doi.org/10.1016/j.enggeo.2011.09.006.

Nikoobakht S., Azarafza M., Akgün H., Derakhshani R. (2022). Landslide susceptibility assessment by using convolutional neural network. *Applied Sciences*, 12(12), 5992. https://doi.org/10.3390/app12125992.

Pourghasemi H.R., Yansari, Z.T., Panagos P., Pradhan B. (2018). Analysis and evaluation of landslide susceptibility: a review on articles published during 2005–2016 (periods of 2005–2012 and 2013–2016). *Arabian Journal of Geosciences*, 11, 1-12. https://doi.org/10.1007/s12517-018-3531-5.

Reichenbach P., Rossi M., Malamud B.D., Mihir M., Guzzetti F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, 180, 60-91. https://doi.org/10.1016/j.earscirev.2018.03.001.

Süzen M.L., Doyuran V. (2004). A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate. *Environmental Geology*, 45, 665-679. https://doi.org/10.1007/s00254-003-0917-8.

Zêzere J.L. (2002). Landslide susceptibility assessment considering landslide typology. A case study in the area north of Lisbon (Portugal). *Natural Hazards and Earth System Sciences*, 2(1/2), 73-82. https://doi.org/10.5194/nhess-2-73-2002.