Machine Learning Approach to Identify Promising Mountain Hiking Destinations Using GIS and Remote Sensing

Lahbib Naimi[1], Charaf Ouaddi[2], Lamya Benaddi[3], El Mahi Bouziane[4], Abdeslam Jakimi[5], Mohamed Manaouch[6]
Software Engineering and Information Systems Engineering Team, Department of Informatics,
Faculty of Sciences and Techniques of Errachidia, Moulay Smail University, Errachidia 52000, Morocco[1,2,3,4,5]
Department of Geography, Faculty of Humanities and Social Sciences,
Ibn Tofail University, Kenitra 14000, Morocco[6]

***Abstract*—The objective of this study is to address the complex task of identifying optimal locations for mountain hiking sites in the Eastern High Atlas region of Morocco, considering topographical factors. The study assesses the effectiveness of a commonly used machine learning classifier (MLC) in mapping potential mountain hiking areas, which is crucial for promoting and enhancing tourism in the area. To begin with, an extensive inventory of 120 mountain hiking sites was conducted, and precise measurements of three topographical parameters were collected at each site. Subsequently, a machine learning algorithm called Bagging was employed to develop a predictive model. The model achieved a high performance, with an area under the curve (AUC) value of 0.93. The model effectively identified favorable areas, encompassing around 24% of the study region, which were predominantly located in the western part. These areas were characterized by mountainous terrain, shorter slopes, and higher altitudes. The research findings provide valuable guidance to decision-makers, offering a roadmap to enhance the discovery of mountain hiking sites in the region.**

Keywords*—Machine learning; mountain hiking; AI-based tourism; GIS; remote sensing; tourism; bagging algorithm; decision-making

I. Introduction

Mountain hiking, also known as mountain trekking or mountaineering, is an outdoor recreational activity that involves exploring and traversing mountainous regions on foot. It typically involves escalating and descending mountains or traversing steep and rugged terrain. Mountain hiking often requires physical endurance, technical skills, and knowledge of mountaineering techniques such as navigation, rock climbing, and rope handling. Hikers engage in this activity to experience the natural beauty of mountain landscapes, challenge themselves physically and mentally, and achieve personal goals. Mountain hiking can range from leisurely day hikes on well-marked trails to multi-day expeditions that may involve camping and navigating through remote and challenging environments.

For this purpose, mountaineering has been addressed in several studies regarding its relationship with tourism, the environment, and the emotional aspects of mountaineers. The study conducted by Wang et al. [1] focuses on the contribution of mountaineering to the creation of lasting and impactful tourism experiences. By utilizing Mount Huangshan in China as a case study, the research explores how sensory perceptions, physical engagement, and emotional connections play a pivotal role in shaping the memories of tourists who engage in hiking activities within this renowned mountainous region.

On the other hand, the research conducted by Galiakbarov et al. [2] delved into the psychological and spiritual aspects of mountaineering, using calling theory to explore why individuals engage in this activity. Their research sheds light on the deeply personal motivations behind mountaineering as a purposeful and transformative experience. However, the study lacks focus on geographic and environmental factors that influence the suitability of certain regions for mountaineering, creating a gap that this research aims to address.

Several studies have also investigated the environmental aspects of mountaineering. Naseri et al. [3] examined the environmental, social, and economic dimensions of sustainable development in mountaineering tourism. Their work emphasized the need for sustainable practices to protect fragile mountain ecosystems while fostering tourism. However, geographic-specific methodologies for identifying suitable hiking sites, particularly in under-researched areas like the High Atlas of Morocco, remain limited. This gap becomes especially relevant in regions with rich biodiversity and geological formations, where tourism can play a pivotal role in both economic development and environmental conservation.

Liu et al. [4] explored the connection between leisure participation in mountaineering and environmentally responsible behavior, focusing on how emotional bonds with landscapes (place attachments) influence conservation attitudes. This emphasizes the importance of understanding not only the environmental but also the cultural and emotional factors that drive sustainable tourism development.

Similarly, Rogerson et al. [5] traced the historical evolution of mountaineering tourism in South Africa, showing how cultural and historical significance contributes to the development of special interest tourism. While these studies offer valuable insights, they do not address how modern technological tools like machine learning algorithms can assist in identifying new areas for mountain tourism based on geographical and topographical factors.

In Morocco, research on mountain hiking tourism remains scarce, particularly in the eastern High Atlas region. A notable exception is the study by Kchikach et al. [6], which focuses on the UNESCO M'Goun Geopark in the central High Atlas. The eastern High Atlas, however, despite being known for its jagged peaks, deep gorges, and ancient rock structures, has not been extensively studied in the context of tourism development. This region holds significant potential for mountain hiking and climbing due to its unique geological formations and natural beauty, making it an ideal candidate for tourism-related research.

To address this gap, this study introduces a novel approach by leveraging machine learning algorithms, specifically Bagging, to identify potential hiking sites in the eastern High Atlas. Bagging, a robust ensemble learning method, was selected due to its ability to improve the accuracy and stability of predictions, particularly in geographic modeling tasks where data variability is high.

Furthermore, the study uses the Area Under the Curve (AUC) and Receiver Operating Characteristic (ROC) metrics to evaluate model performance, ensuring a rigorous assessment of the results. These methods are widely accepted for evaluating the accuracy of predictive models, particularly in machine learning-based geographic studies. The combination of AUC and ROC analysis with Bagging ensures a comprehensive approach to identifying suitable mountain hiking locations based on geographic data.

In addition to the technical aspects, this research has broader implications for sustainable tourism development in Morocco. By identifying suitable hiking locations, the study contributes to the sustainable management of natural resources and helps promote ecotourism in underdeveloped areas. Sustainable tourism practices, which are essential for preserving the environmental and cultural heritage of the region, can be informed by the findings of this study. As tourism continues to grow in the High Atlas, ensuring that it is developed in an environmentally responsible manner is crucial for maintaining the long-term viability of these destinations.

The implications of this research extend beyond the immediate identification of suitable hiking sites. By integrating advanced machine learning techniques into the tourism development process, this study establishes a precedent for leveraging technology to advance the tourism sector in Morocco and globally [7], [8], while also enhancing environmental conservation efforts.

The findings will not only support sustainable tourism initiatives but also contribute to the long-term preservation of the High Atlas's unique ecosystems and cultural heritage. This alignment of tourism with conservation principles can foster a more responsible tourism industry, encouraging both local communities and policymakers to prioritize sustainable practices that benefit the environment and the economy

II. Materials and Methods

A. Study Area

The upper Ziz area in the eastern High Atlas region of southeastern Morocco (Fig. 1) experiences a semi-arid climate with cold winters and dry summers, as indicated by Manaouch et al. [9]. The average annual precipitation ranges from 119 to 377 mm, and temperatures vary between 10.2 and 19.2°C, as documented by Manaouch et al. [10]. The elevation ranges from 1023 to 3687 meters above sea level, with slopes ranging from 0 to 66°C, as noted by Manaouch et al. [11].

This region boasts numerous mountain hiking sites that showcase its unique geological and geomorphological features. These sites provide opportunities for tourism, offering insights into the region's geological history. Among these sites, the Gorges of Ziz are particularly remarkable. Carved by the Ziz river over millennia, these limestone gorges feature towering cliffs and winding canyons, creating a breathtaking landscape. Their geological significance and picturesque beauty make them a magnet for geologists and nature enthusiasts alike.

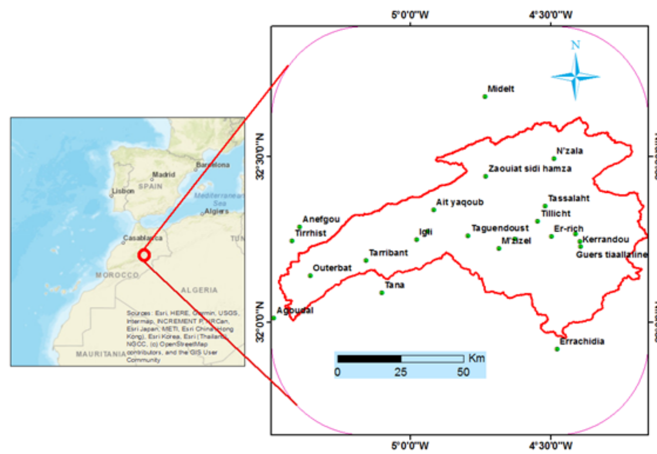

Fig. 1. Location of the eastern high atlas of upper Ziz in Morocco.

B. Data Used and Mountain Hiking Sites Inventory

Accurate identification and selection of mountain hiking sites require a thorough analysis of the phenomenon and its underlying factors [12]. This study conducted extensive field visits in the eastern High Atlas of upper Ziz, SE Morocco, identifying and modeling a total of 120 mountain hiking sites. An equal number of non-mountain hiking sites were also considered to ensure a balanced approach (Fig. 2).

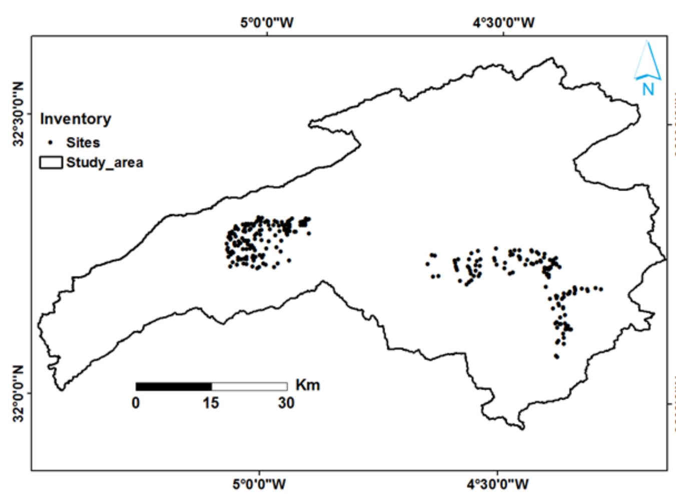

Fig. 2. Mountain hiking sites inventory map.

The study incorporated three conditioning factors (slope, length of slope, and elevation) as independent variables influencing the suitability of areas for mountain hiking sites. These variables were measured for the 240 selected sites. The dataset was divided into training data (70%) and validation data (30%) following established spatial modeling conventions.

Data processing was performed using ArcGIS 10.5 and SPSS Statistics 26 software, which enabled the transformation of geospatial data into a tabular format. The dataset was prepared for analysis by importing data, defining attributes, executing geoprocessing tools, configuring parameters, and performing geoprocessing operations. Model validation was conducted using the Area Under the Curve (AUC) of the receiver operating characteristic (ROC), a common method in geospatial modeling studies to assess and validate model reliability [10].

C. Mountain Hiking Sites (MHS)

Mountain hiking sites (MHS) are of paramount importance in various fields such as environmental education, recreation, and sports [13]. These sites not only offer opportunities for exploration and the appreciation of natural landscapes but also serve as key locations for ecological studies, conservation efforts, and physical fitness activities [14]. Canyons, cliffs, and other geological formations are significant not just for their scenic beauty but also for the insight they provide into the Earth's geological history, making them valuable for scientific research and outdoor education [15].

In addition to their value in education and conservation, MHS contribute to local economies by supporting activities such as mountaineering, hiking, and outdoor excursions, which often require the development of infrastructure such as trails, shelters, and conservation measures. These efforts foster sustainable practices and environmental stewardship while providing recreational opportunities [16].

Fig. 3. Mountainous hiking sites and mountain hikers encountered in the upper Ziz.

While mountain hiking sites are frequently associated with tourism, their broader significance spans disciplines like environmental science, geography, and physical education [17]. In particular, these sites are ideal for studying geomorphological processes, fostering a deeper understanding of natural erosion and sedimentation patterns [18]. Mountain hiking activities, especially in regions like the upper Ziz valley, support not only recreation but also health and well-being through physical fitness and mental restoration [19].

Fig. 3 provides visual examples of mountain hiking sites and hikers along the Ziz valley, showcasing their versatility and value across multiple domains.

III. Methodology

Fig. 4 depicts a flowchart that outlines the sequential procedure employed to evaluate the suitability of mountain hiking locations utilizing an advanced machine learning algorithm. The methodology begins with gathering data from diverse sources, including Google Earth images, field surveys, geological maps, Landsat 8 OLI images, and previous studies. By consolidating this information, a comprehensive database is created, serving as the foundation for the analysis.

Key mountain hiking conditioning factors (MHCFs) such as slope, length of slope, and elevation are identified and incorporated into the study. These factors are crucial in determining the suitability of hiking locations. An inventory map is developed, detailing 240 sites that serve as reference points for the evaluation process. The MHCFs undergo geoprocessing, which involves manipulating spatial data to derive meaningful insights. This step ensures that the data is in a suitable format for further analysis. The processed data is stored in a centralized database, facilitating convenient access and management.

To build and validate the machine learning model, the data is split into training and testing sets, with 70% allocated for training and 30% for testing. This division is essential for developing a robust and reliable model. The process also includes calculating frequency ratios, normalizing data, examining spatial relationships, and preparing data layers, all of which ensure the data is ready for analysis.

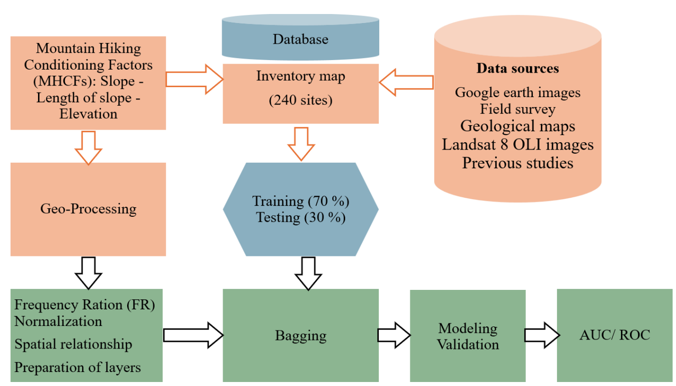

Fig. 4. Flowchart of the modeling strategy used.

The final step involves evaluating the model's performance using AUC (Area Under the Curve) and ROC (Receiver Operating Characteristic) metrics. These metrics provide a measure of the model's accuracy and effectiveness, ensuring the methodology yields reliable results for assessing mountain hiking locations. The subsequent sections offer a comprehensive and detailed explanation of this methodology, providing a deeper understanding of the approach.

A. *Mountain Hiking Conditioning Factors (MHCFs)*

Topographic factors such as slope, slope length, and elevation significantly impact the suitability of mountainous areas for hiking. These factors influence the characteristics and spatial distribution of land formations. The following sections discuss each factor in detail and provide visual maps representing them.

1) Elevation: Elevation is a crucial attribute in mountain hiking site identification as it helps differentiate various landforms. Mountain hiking sites may vary significantly in elevation, influencing the climate, vegetation, and geological processes of the area. Surveying techniques, satellite-based systems (e.g. GPS), and remote sensing methods like LiDAR or radar are commonly used to measure elevation. Fig. 5 shows the elevation map of the study area.

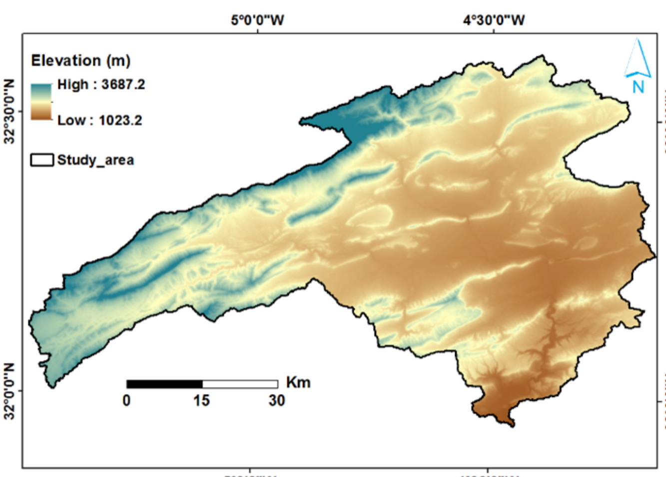

Fig. 5. MHCF map: Elevation.

2) Slope: Slope refers to the steepness or inclination of a landform's surface, a key factor in evaluating hiking site suitability. The slope is calculated in ArcGIS using the variation in elevation between adjacent cells within a Digital Elevation Model (DEM) raster dataset. The formula for calculating slope is:

$$\text{Slope } (\%) = \left(\frac{\text{rise}}{\text{run}}\right) \times 100 \tag{1}$$

Where:

- *rise* represents the change in elevation between cells.
- *run* represents the horizontal distance between the cells.

Fig. 6 illustrates the slope map derived from the DEM.

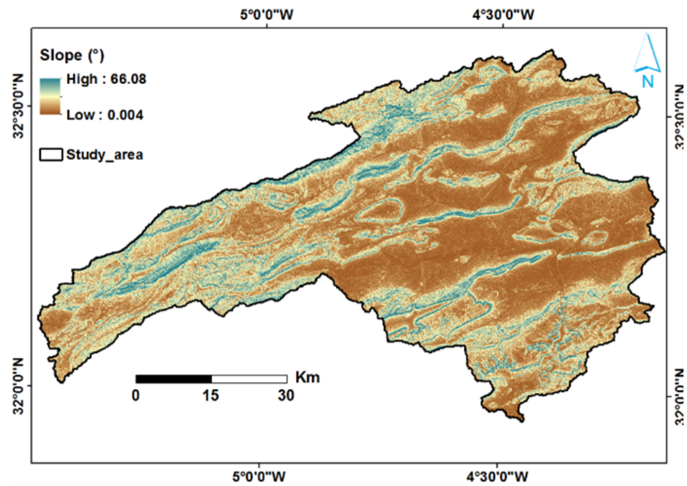

Fig. 6. MHCF map: Slope.

3) Length of Slope (LS): The length of the slope (LS) helps characterize landforms within mountain hiking sites. LS is derived from the DEM using ArcGIS 10.5 software based on the methodology developed by Bizuwerk et al. [20], following Wischmeier and Smith's formula:

$$\text{LS} = \left(\frac{L}{22.13}\right)^m \times \left(65.41\sin^2(S) + 4.56\sin(S) + 0.065\right) \tag{2}$$

where,

- L is the slope length.
- S is the slope gradient (expressed as a percentage).
- m is a constant value that varies depending on the range of slope gradients.

The resulting LS map is shown in Fig. 7.

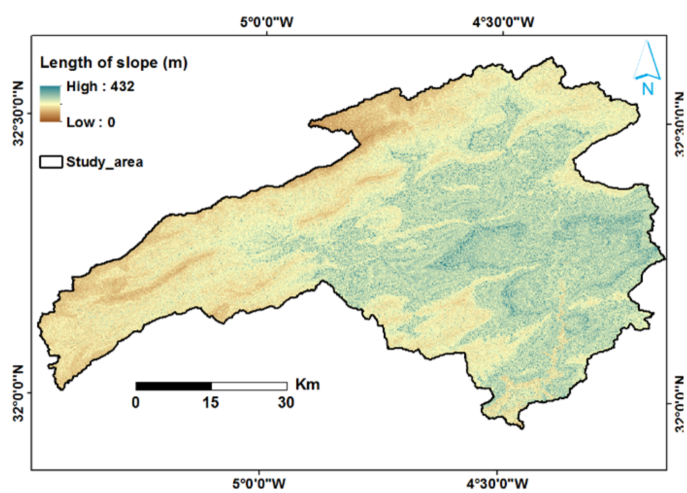

Fig. 7. MHCF map: Length of slope.

B. Machine Learning Algorithm and Frequency Ratio (FR)

1) Bagging algorithm: The Bagging algorithm (Bootstrap Aggregating) comprises two main steps, as described by Breiman et al. [21]. In the first step, multiple training datasets are created by resampling from the original data, a process known as bootstrapping. This involves randomly selecting observations from the original dataset with replacement, allowing some observations to be repeated while others may be omitted. This resampling technique helps to reduce variance and improve the stability of the model.

In the second step, multiple models are constructed using these bootstrapped training datasets. Each model is trained independently, allowing them to capture different aspects of the data. The final predictions are then obtained by aggregating the results of these individual models, typically through averaging for regression tasks or majority voting for classification tasks. This aggregation process tends to produce a more robust and accurate model than any single constituent model.

Bagging has been widely applied in geospatial modeling tasks, where the inherent variability in spatial data can significantly impact model performance. For instance, studies like [22] and [23] have demonstrated the effectiveness of Bagging in enhancing predictive accuracy for various geospatial phenomena.

This model was chosen for this study due to its ability to reduce variance and enhance model stability, particularly in geographic modeling tasks, where data variability can significantly impact predictive performance. While Random Forest is another popular ensemble method frequently applied in geospatial analysis [24], [25], it is a variant of Bagging that focuses on constructing multiple decision trees. In contrast, Bagging can be applied with a variety of base learners, offering greater flexibility. Random Forest incorporates random feature selection at each split, which can be advantageous for certain types of data but may lead to higher bias when modeling complex spatial patterns.

Support Vector Machines (SVM) are also widely used in geographic information system (GIS) modeling due to their strong performance in high-dimensional spaces [26]. However, SVM can be sensitive to noise in the data and may require extensive parameter tuning, especially in cases where the data distribution is complex and non-linear. Bagging, by aggregating the predictions of multiple base models, mitigates the risk of overfitting and noise sensitivity, providing a more balanced approach to spatial data modeling.

2) Frequency Ratio (FR): The goal of analyzing the mountain hiking conditioning factors is to carefully assess how different geographic parameters relate to the presence of mountain hiking sites. To do this, we use the frequency ratio method, which allows us to measure how often hiking sites appear in different classes of geographic factors (such as elevation, slope, or land use) compared to how much area those classes cover. As explained by Manaouch et al. [23], this method helps determine the likelihood of finding a hiking site in specific geographic conditions.

The frequency ratio technique works by comparing two percentages: the percentage of mountain hiking sites located in a certain class and the percentage of the total area that this class occupies. A higher frequency ratio means there is a greater chance of finding a hiking site in that specific geographic class, as noted by Samanta et al. [27].

The frequency ratio for a particular class is calculated using the following equation:

$$FR_i = \frac{\left(\frac{N_i}{N}\right)}{\left(\frac{S_i}{S}\right)} \tag{3}$$

Where:

- N_i is the number of mountain hiking sites within the ith class of a geographic factor (e.g., a specific elevation range or slope category).
- N is the total number of mountain hiking sites in the study area.
- S_i is the area of the ith class of the geographic factor.
- S is the total area of the study region.

This formula helps us understand whether certain geographic conditions are more favorable for mountain hiking. For example, if a specific elevation range covers 20% of the study area but contains 50% of the hiking sites, the frequency ratio will be higher, indicating that this elevation is particularly suitable for hiking.

The frequency ratio equation is used to compute FR_i for each class of a conditioning factor, helping us evaluate how well different geographic conditions align with the locations of hiking sites.

C. Performance and Model Validation

The model's performance was evaluated using the Area Under the Curve (AUC) of the receiver operating characteristic (ROC) analysis. This analysis provides a comprehensive measure of the model's ability to distinguish between suitable and unsuitable mountain hiking areas. The AUC value ranges from 0 to 1, where a value of 0.5 indicates no discriminative power and a value of 1 signifies perfect discrimination.

For this study, the AUC/ROC analysis involved 120 mountain hiking sites and the classified map of potential hiking areas. True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) rates were calculated, which serve as critical metrics in evaluating the model's predictive accuracy. These metrics were derived using the following equations:

$$TPR = \frac{TP}{TP + FN} \tag{4}$$

$$FPR = \frac{FP}{FP + TN} \tag{5}$$

In this context, the True Positive Rate (TPR), also known as sensitivity or recall, reflects the proportion of actual positives that were correctly identified by the model. Conversely, the False Positive Rate (FPR) quantifies the proportion of actual negatives that were incorrectly classified as positives. By analyzing these rates, we can better understand the trade-offs between sensitivity and specificity, which is crucial for applications where the consequences of false positives and false negatives differ significantly.

TABLE I. Data Labeling for ROC Analysis

Actual Label	Predicted Positive	Predicted Negative
Positive	TP	FN
Negative	FP	TN

Table I summarizes the data labeling used for the ROC analysis, indicating how actual labels correspond to predicted outcomes. This structured approach enables a clearer evaluation of the model's performance across different thresholds.

The resulting AUC value of approximately 0.93 indicates that the Bagging classifier performed exceptionally well in accurately identifying and distinguishing potential mountain hiking sites. Such validation not only affirms the reliability of the model but also enhances confidence among stakeholders in utilizing these findings for informed decision-making regarding tourism development.

Furthermore, the ROC curve, which graphically represents the TPR against the FPR at various threshold settings, provides an intuitive visualization of the model's performance. Analyzing the shape and area under this curve can reveal insights into the model's robustness, assisting in the selection of the optimal threshold that balances sensitivity and specificity according to the specific context of mountain hiking site selection.

D. Mountain Hiking Sites and Conditioning Factors

Table II presents the calculated frequency ratio values for the three conditioning factors, each consisting of multiple classes. To facilitate analysis, these values have been normalized using the normalize filter method, which transforms them into a range of 0.1 to 0.9. A normalized value closer to 1 indicates a stronger association between mountain hiking sites and the corresponding factor, while a value closer to 0 indicates a weaker association. The spatial relationship between the conditioning factors and mountain hiking sites is assessed through the frequency ratio calculation, and the results are presented in Table II for all factors.

E. Mountain Hiking Suitability Map

Fig. 8 illustrates the mountain hiking suitability map generated by the Bagging classifier. The map highlights that the northern and western regions exhibit higher suitability for potential mountain hiking sites compared to other areas. To gain a deeper understanding of the factors influencing the suitability or unsuitability of different areas and to explore the underlying reasons for these variations, we collaborated with local experts who possess substantial knowledge of the region.

Additionally, we validated our findings by cross-referencing them with images obtained from Google Earth. Fig. 10 provides the corresponding percentages of the area covered by each suitability class in the generated map.

TABLE II. Spatial Relationship between Slope and Mountain Climbing Sites Points using Frequency Ratio (FR)

RCF	Class	Frequency Ratio FRi	Frequency Ratio FRn
Slope	1	0.045	0.1
	2	0.3	0.12
	3	0.74	0.15
	4	1.43	0.21
	5	9.58	0.9
Elevation	1	0.1	0.1
	2	0.2	0.1
	3	0.69	0.13
	4	0	0.1
	5	0.2	0.1
	6	0	0.1
	7	7.6	0.55
	8	4.17	0.34
	9	0	0.1
Length of slope	1	0	0.1
	2	0	0.1
	3	0.16	0.11
	4	0.5	0.24
	5	2.02	0.81

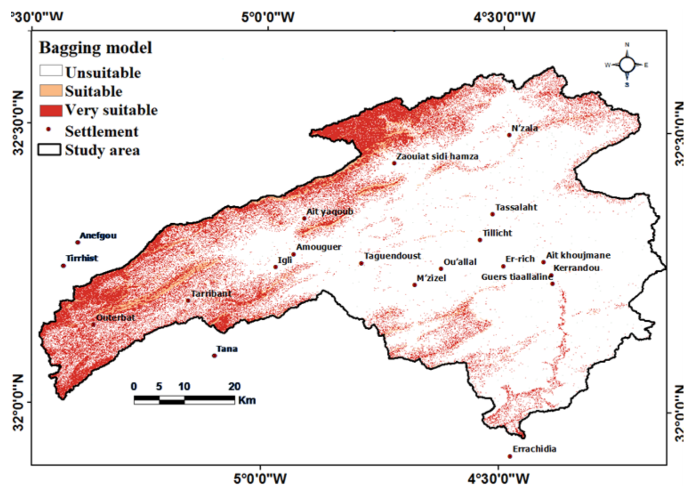

Fig. 8. Generated mountain hiking suitability map using Bagging.

F. Model's Performance

The effectiveness of the Bagging classifier was evaluated using the AUC/ROC method, which is widely recognized for its robustness in assessing model performance in binary classification tasks. During the training phase, the classifier demonstrated impressive performance, achieving an average accuracy rate of 97.29%. This high accuracy reflects the model's capability to learn from the training data effectively and underscores its potential in identifying suitable mountain hiking sites.

However, during the validation phase, significant fluctuations in the classifier's performance were observed, as depicted in Fig. 9. These fluctuations can be attributed to several factors, including variability in the validation dataset, potential overfitting during the training phase, and the inherent complexity of the mountain hiking suitability classification task. Understanding these fluctuations is crucial, as they highlight the necessity for careful model tuning and validation to ensure reliable outcomes in real-world applications.

The obtained AUC value of 0.935 indicates that the Bagging classifier successfully modeled the suitability mapping of mountain hiking sites in the upper Ziz region. An AUC value above 0.9 is generally considered excellent, suggesting that the model has a high probability of distinguishing between suitable and unsuitable hiking areas. This level of performance is particularly encouraging, as it implies that the classifier can reliably inform decisions regarding the identification of potential hiking locations.

Fig. 9 illustrates the prediction rate curve for the mountain hiking suitability map produced by the Bagging classifier. The curve visually represents the relationship between the true positive rate (TPR) and the false positive rate (FPR) across various threshold settings. Analyzing this curve allows for a deeper understanding of the trade-offs involved in selecting an optimal threshold, ensuring that the final suitability map meets the specific objectives of identifying suitable hiking sites.

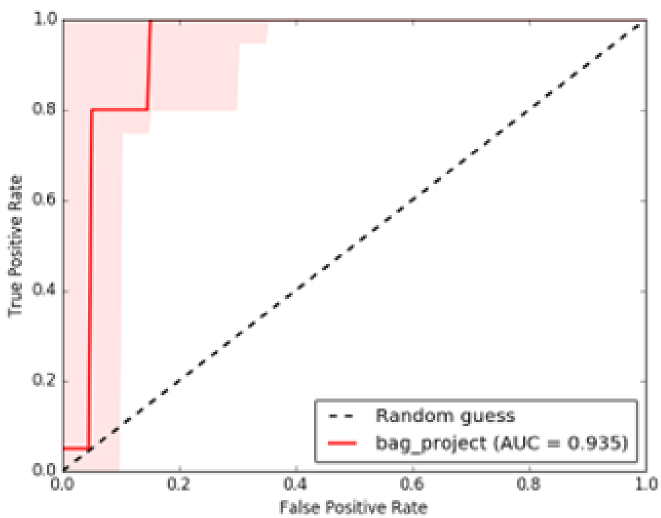

Fig. 9. Prediction rate curve for mountain hiking suitability map produced by Bagging.

IV. Discussion

A. Analysis of Precision

The performance of machine learning models (MLMs) can vary depending on the specific algorithm employed. In the case of Bagging, its performance was assessed using the 10-fold cross-validation method during the training phase, resulting in a high level of performance with an AUC value of 0.97.

However, during the validation phase, the performance of Bagging decreased to 93%, as depicted in Fig. 9. It is generally observed that algorithms with higher AUC values tend to exhibit more accurate and efficient prediction capabilities, as emphasized by Su et al. [28]. Considering the current AUC value of Bagging, it can be inferred that Bagging performs relatively lower compared to the results reported by Manaouch et al. [23] in predicting potential reforestation areas. Nevertheless, it is important to note that Bagging still demonstrates a favorable level of accuracy and efficiency, as evidenced by its AUC value of 0.93.

In evaluating the model's accuracy, it is essential to consider other factors that influence the suitability of regions for mountain hiking beyond the quantitative metrics provided by the model. For instance, cultural aspects, such as local traditions and community engagement in outdoor activities, can significantly impact the attractiveness of certain hiking locations. Furthermore, the presence of infrastructure, including access roads, trail maintenance, and proximity to amenities (like parking and rest areas), plays a crucial role in determining the feasibility and safety of hiking sites. Areas that may have high ecological or aesthetic value but lack necessary infrastructure may be less suitable for hiking compared to less pristine areas that are well-supported by facilities.

B. Distribution of Suitable Mountainous Hiking Areas

The data attributes collected from the upper Ziz region underwent preprocessing using the pre-trained Bagging classifier. During this preprocessing stage, mountain hiking suitability indices were computed for each pixel within the study area. The "Natural breaks" algorithm was employed to classify these suitability indices, resulting in the creation of three distinct classes: unsuitable, suitable, and very suitable. This process led to the generation of a mountain hiking suitability map, as depicted in Fig. 8.

The findings indicate that approximately 24% of the entire study area comprises very suitable areas for potential mountain hiking sites, as shown in Fig. 10. According to the Bagging model, the southern parts of the region are deemed unsuitable for mountain hiking sites, while the mountainous regions situated to the west and north exhibit a high level of suitability. Additionally, the Bagging model reveals that areas classified as very suitable for mountain hiking sites are dispersed along the Ziz wadi, located to the south of the upper Ziz region.

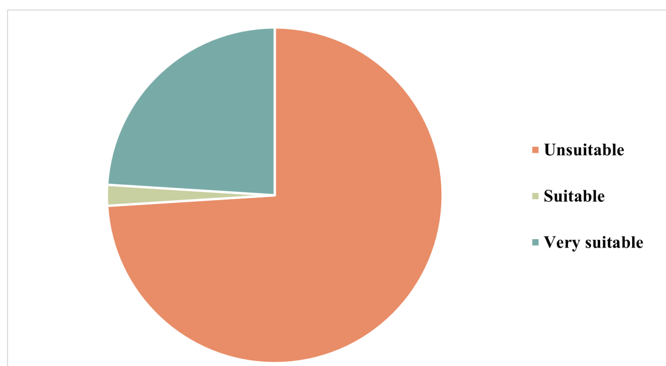

Fig. 10. Percentage of mountain hiking suitability classes' areas for Bagging model.

Despite these promising results, there are several limitations to this study that must be acknowledged. One significant limitation is the potential quality of the data used in the analysis. Incomplete or inaccurate data could adversely affect the model's predictive capabilities. Furthermore, the scope of topographical features included in the analysis may not encompass all relevant variables that could influence hiking suitability, such as microclimate variations, vegetation types, or the impact of human activities on the landscape. A broader dataset could enhance the model's accuracy and provide a more comprehensive understanding of the factors affecting hiking suitability.

V. Conclusion and Future Work

The integration of topographical data and the Bagging algorithm has demonstrated significant potential in enhancing tourism through the identification of suitable mountain hiking sites. By utilizing the Bagging classifier, which incorporates critical topographical features such as slope, length of slope, and elevation, and validating the resulting map using the AUC/ROC metric, valuable insights can be gained to identify potential mountain hiking sites in the eastern High Atlas of the upper Ziz in Southeast Morocco. The findings indicate that approximately 24% of the study area comprises very suitable areas for potential mountain hiking sites, primarily located in the northern and western regions, as well as around the Ziz wadi, particularly in the downstream part of the study area.

The application of the AUC/ROC evaluation metric in this study provides a robust measure of the algorithm's ability to differentiate between suitable and unsuitable mountain hiking sites, achieving an impressive score of approximately 93%. This indicates that the Bagging algorithm performed exceptionally well in accurately identifying and distinguishing potential mountain hiking sites. Such validation approaches not only ensure the reliability and accuracy of the generated map but also empower decision-makers and stakeholders in the tourism industry to confidently identify and prioritize areas for tourism development.

Furthermore, the findings of this study underscore the importance of employing machine learning techniques in environmental and tourism planning. By systematically analyzing the spatial relationships among various topographical factors, stakeholders can make informed decisions that promote sustainable tourism practices while preserving natural landscapes.

Future research could explore the integration of additional ecological and socio-economic data to refine the suitability assessments further. Incorporating factors such as ecological diversity and human accessibility, including proximity to transport links and local amenities, can significantly enhance the accuracy of suitability models. Additionally, collaborating with local communities is crucial for validating findings. Engaging with residents can provide insights into cultural significance, traditional knowledge, and existing infrastructure, which are essential for a comprehensive understanding of the area's tourism potential.

Moreover, applying similar methodologies in other geographical areas could contribute to a broader understanding of potential hiking sites, ultimately supporting the growth of ecotourism on a larger scale. Future studies should also consider the impacts of climate change on hiking site suitability, ensuring that long-term strategies for tourism development are both adaptable and sustainable.

Overall, this study not only contributes to the literature on tourism but also offers practical applications for enhancing the visitor experience in mountain environments, fostering both environmental conservation and local economic development.

Acknowledgment

This research project was supported by funding from the CNRST (National Center for Scientific and Technical Research) under the Moroccan project titled "Al-Khawarizmi program in AI and its Applications", (Alkhawarizmi/2020/32).

The authors would like to express their gratitude to Mohamed Manaouch for providing the necessary facilities and data for this study. Additionally, the authors would like to extend their thanks to Mr. N. Khan for his invaluable assistance during the validation phase

References

[1] Y. Wang, L. Zhou, H. Chen, L. Wang, and X. Wu, "Framing memorable hiking tourism experiences through embodiment: The case of mount huangshan, china," *Journal of Outdoor Recreation and Tourism*, vol. 45, p. 100710, 2024.

[2] Y. Galiakbarov, O. Mazbayev, L. Mutaliyeva, V. Filimonau, and H. Sezerel, "When the mountains call: Exploring mountaineering motivations through the lens of the calling theory," *Journal of Outdoor Recreation and Tourism*, vol. 45, p. 100743, 2024.

[3] H. Naseri, P. Nourbakhsh, M. Kohandel, and A. Khodayari, "Factors affecting sustainable development in mountaineering," *Environmental Education and Sustainable Development*, vol. 12, no. 2, pp. 165–183, 2024.

[4] Z. Liu, T. Yang, C. Yi, and K. Zhang, "Effects and functional mechanisms of serious leisure on environmentally responsible behavior of mountain hikers," *Journal of Outdoor Recreation and Tourism*, vol. 45, p. 100709, 2024.

[5] C. M. Rogerson and J. M. Rogerson, "Historical special interest tourism: The evolution of mountaineering in south africa," *Revistă Româna de Geografie Politică*, vol. 26, no. 1, pp. 1–13, 2024.

[6] N. Kchikach, H. Ibouh, A. Benali, A. Charbaoui, and A. E. A. El Fels, "Geoheritage of a mountain tourist route (high atlas, morocco): Digital promotion and visitor typology," 2024.

[7] L. Naimi, E. M. Bouziane, and A. Jakimi, "A model-driven approach for developing smart tourism web applications," in *International Conference on Digital Technologies and Applications*. Springer, 2024, pp. 255–263.

[8] L. Naimi, A. Jakimi, L. El Bermi *et al.*, "A sketch of a mda approach to accelerate the development of user interfaces in tourism web applications," in *2024 Mediterranean Smart Cities Conference (MSCC)*. IEEE, 2024, pp. 1–5.

[9] M. Manaouch, S. Mohamed, and F. Imad, "Coupling fuzzy logic and analytical hierarchy process (fahp) with gis for landslide susceptibility mapping (lsm) in ziz upper watershed, se morocco," 2021.

[10] M. Manaouch, M. Sadiki, M. Aghad, Q. B. Pham, M. Batchi, and J. Al Karkouri, "Assessment of landslide susceptibility using machine learning classifiers in ziz upper watershed, se morocco," *Physical Geography*, vol. 45, no. 2, pp. 203–230, 2024.

[11] M. Manaouch, M. Sadiki, and I. Fenjiro, "Integrating gis-based fahp and watem/sedem for identifying potential rwh areas in semi-arid areas," *Geocarto International*, vol. 37, no. 25, pp. 8882–8905, 2022.

[12] A. M. Al-Abadi, "Mapping flood susceptibility in an arid region of southern iraq using ensemble machine learning classifiers: a comparative study," *Arab. J. Geosci*, vol. 11, p. 218, 2018.

[13] P. Beedie and S. Hudson, "Emergence of mountain-based adventure tourism," *Annals of Tourism Research*, vol. 30, no. 3, pp. 625–643, 2003.

[14] R. Buckley, "Impacts of ecotourism on birds." 2004.

[15] G. Herrera-Franco, P. Carrión-Mero, N. Montalván-Burbano, J. Caicedo-Potosí, and E. Berrezueta, "Geoheritage and geosites: A bibliometric analysis and literature review," *Geosciences*, vol. 12, no. 4, p. 169, 2022.

[16] D. B. Weaver, "Ecotourism as mass tourism: Contradiction or reality?" *Cornell Hotel and Restaurant Administration Quarterly*, vol. 42, no. 2, pp. 104–112, 2001.

[17] K. Thapa, D. King, Z. Banhalmi-Zakar, and A. Diedrich, "Nature-based tourism in protected areas: a systematic review of socio-economic benefits and costs to local people," *International Journal of Sustainable Development & World Ecology*, vol. 29, no. 7, pp. 625–640, 2022.

[18] R. Huggett and E. Shuttleworth, *Fundamentals of Geomorphology*. Routledge, 2022.

[19] D. Huber, J. Freidl, C. Pichler, M. Bischof, M. Kiem, R. Weisböck-Erdheim, G. Squarra, V. De Nigris, S. Resnyak, M. Neberich *et al.*, "Long-term effects of mountain hiking vs. forest therapy on physical and mental health of couples: a randomized controlled trial," *International journal of environmental research and public health*, vol. 20, no. 2, p. 1469, 2023.

[20] A. Bizuwerk, G. Taddese, and Y. Getahun, "Application of gis for modelling soil loss rate in awash river basin," *Ethiopia, ILRI*, 2003.

[21] L. Breiman, "Using iterated bagging to debias regressions," *Machine Learning*, vol. 45, pp. 261–277, 2001.

[22] Q. Zhang, Z. Ning, X. Ding, J. Wu, Z. Wang, P. Tsangaratos, I. Ilia, Y. Wang, and W. Chen, "Hybrid integration of bagging and decision tree algorithms for landslide susceptibility mapping," *Water*, vol. 16, no. 5, p. 657, 2024.

[23] M. Manaouch, M. Sadiki, Q. B. Pham, A. Zouagui, M. Batchi, and J. Al Karkouri, "Predicting potential reforestation areas by quercus ilex (l.) species using machine learning algorithms: case of upper ziz, southeastern morocco," *Environmental Monitoring and Assessment*, vol. 195, no. 9, p. 1094, 2023.

[24] M. Geerts, S. vanden Broucke, and J. De Weerdt, "Georf: a geospatial random forest," *Data Mining and Knowledge Discovery*, pp. 1–35, 2024.

[25] H. Talebi, L. J. Peeters, A. Otto, and R. Tolosana-Delgado, "A truly spatial random forests algorithm for geoscience data analysis and modelling," *Mathematical Geosciences*, vol. 54, no. 1, pp. 1–22, 2022.

[26] K. Sun, Z. Li, S. Wang, and R. Hu, "A support vector machine model of landslide susceptibility mapping based on hyperparameter optimization using the bayesian algorithm: a case study of the highways in the southern qinghai–tibet plateau," *Natural Hazards*, pp. 1–22, 2024.

[27] S. Samanta, D. K. Pal, and B. Palsamanta, "Flood susceptibility analysis through remote sensing, gis and frequency ratio model," *Applied Water Science*, vol. 8, no. 2, p. 66, 2018.

[28] Q. Su, W. Tao, S. Mei, X. Zhang, J. Guo, and Y. Yang, "Landslide susceptibility zoning using c5.0 decision tree, random forest, support vector machine and comparison of their performance in a coal mine area," *Front. Earth Sci*, vol. 1181, 2021.