

Segmenting traffic accident risk using remote sensing and GIS data: A comparative analysis of Random Forest and Weighted Overlay Methods

Ha Le Thi^{1,2}, Thao Vu Thi Phuong^{1,*}, Thao Do Thi Phuong¹

¹Hanoi University of Mining and Geology, 18 Vien street, Dong Ngac Ward, Hanoi, Vietnam

²Campus in Ho Chi Minh City, University of Transport and Communications. No.450-451 Le Van Viet Street, Tang Nhon Phu Ward, Ho Chi Minh City, Vietnam

*Email: vuthiphuongthao@humg.edu.vn

Abstract: Identifying traffic accident risk zones is critical for infrastructure management and traffic safety planning. This study compares the performance of two spatial analysis approaches - Weighted Overlay Analysis (WOA) and the Random Forest (RF) machine learning algorithm - in mapping accident risk zones in Thanh Hoa Province, Vietnam. The analysis incorporates four input layers derived from remote sensing and GIS data: Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST), accident density from Kernel Density Estimation (KDE), and Euclidean distance from main roads. In WOA, expert-assigned weights were KDE 30%, LST 20%, NDVI 15%, and Euclidean distance from main roads 35%. In contrast, the RF model is trained on 1,594 actual accident points (2020–2023) and an equal number of non-accident points — showed variable importance values of KDE 38.33%, LST 19.50%, NDVI 14.50%, and Euclidean distance from main roads 27.67%. Results indicate that RF outperforms WOA, achieving 81.07% accuracy and an F1-score of 79.31%. KDE was the most influential variable in the RF model, whereas WOA assigned greater weight to Euclidean distance from main roads. WOA achieved an accuracy of 71.45%. While RF more accurately reflects current accident risk patterns, it still misses some points in sparsely populated or complex terrain. The observed performance gap between RF and WOA, together with their differing variable prioritizations, suggests that integrating expert judgment with machine

learning can enhance the comprehensiveness, objectivity, and reliability of traffic accident risk zoning. Overall, this study underscores the complementary value of combining remote sensing, GIS, and machine learning for proactive traffic safety interventions. The study highlights the complementary value of integrating remote sensing, GIS, and machine learning for proactive traffic safety interventions.

Keywords: Normalized Difference Vegetation Index, Land Surface Temperature, Kernel Density Estimation, Random Forest, Weighted Overlay Analysis

INTRODUCTION

Road traffic accidents pose a significant threat to economic growth and public health in developing countries, including Vietnam. This is largely driven by the rapid increase in motor vehicle usage, limited public awareness of traffic regulations, and infrastructure that fails to keep pace with urban expansion. In regions with complex terrain or deteriorating road conditions, the risk of traffic accidents is further exacerbated—often resulting in severe human and material losses (Le et al., 2024a; Le et al., 2024b; Le KG et al., 2024).

Accurately identifying high-risk traffic zones is essential for targeted interventions in transportation planning and accident prevention. Conventional spatial analysis techniques in GIS such as WOA, Analytic Hierarchy Process, and KDE have been widely

used to integrate risk-related factors like population and accident density (Zheng et al., 2024). However, these methods often rely on subjective expert judgment to assign layer weights, which limits their ability to adapt to real-world accident data (Chen et al., 2011; Eskandari et al., 2007). Nowadays, machine learning algorithms, including Random Forest, are increasingly widely used in spatial risk analysis due to their ability to process complex data and determine the importance of input variables (Farhangi et al., 2021; Wang et al., 2023). However, the application of the WOA method and the RF machine learning algorithm in assessing road traffic accident risk zones has been limited. Additionally, remote sensing images demonstrate their significant ability to monitor changes in environmental factors. They facilitate the collection of information about the potential and status of objects across large areas in a short amount of time. As a result, they have become a crucial data source for numerous recent studies (Vu et al., 2024; Rodriguez-Galiano et al., 2012). Especially in the field of transportation science, numerous studies have utilised remote sensing data as input for machine learning models (Liu et al., 2023; Worachairungreung et al., 2021).

This study utilizes remote sensing data - Landsat 8 images, as input data for the two methods to independently compare their zoning efficiency and the ability to analyze the importance of spatial factors affecting accidents, thereby contributing to the assessment of the feasibility and effectiveness of each model. Thus, the primary objective of the study is to develop a traffic accident risk zoning map using two methods: WOA and RF. At the same time, the study focuses on identifying spatial factors that have a significant impact on accidents, thereby supporting safe traffic planning and early warning of high-risk areas.

Data and methodology

Study area

Thanh Hoa Province is in Vietnam's North Central region, spanning latitudes $19^{\circ}18'$

to $20^{\circ}40'$ North and longitudes $104^{\circ}22'$ to $106^{\circ}04'$ East as Figure 1. Serving as a gateway that connects the North and North Central regions, Thanh Hoa boasts a well-developed transport infrastructure compared to many other provinces in Vietnam. However, the provincial road transport system faces several limitations due to uneven infrastructure across different areas. Mountainous areas often lack roads built to high technical standards, resulting in low load-bearing capacity and reduced traffic safety. These deficiencies hinder transportation efficiency and elevate the risk of traffic accidents—especially during adverse weather or peak traffic periods (Le et al., 2024a).

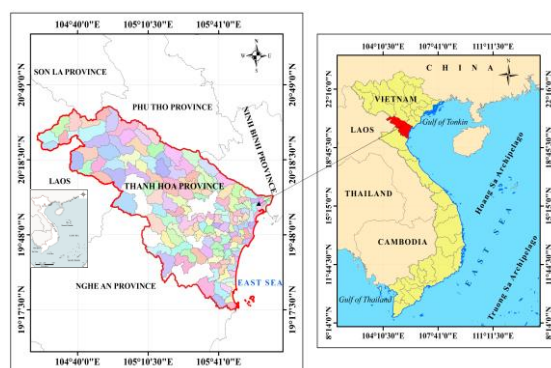


Figure 1. Study area - Thanh Hoa province of Vietnam

Data

The input data for both the WOA and RF models consist of spatial layers derived from remote sensing imagery and field-collected data. Specifically, the Normalized Difference Vegetation Index and Land Surface Temperature were calculated from Landsat 8 satellite images. Traffic accident density was estimated using the KDE method, based on road accident records in Thanh Hoa from 2020 to 2023. The dataset of distances to main roads was created from the provincial traffic network map, in which Euclidean distances were calculated after projecting the data to the WGS_1984_UTM_Zone_48N coordinate system. Additionally, a dataset containing 1,594 accident locations and an equal number of non-accident points was prepared for training

and validating the RF model. The non-accident points were randomly generated within the administrative boundary of Thanh Hoa

Province using the Create Random Points tool in ArcGIS Pro, ensuring no spatial overlap with recorded accident locations.

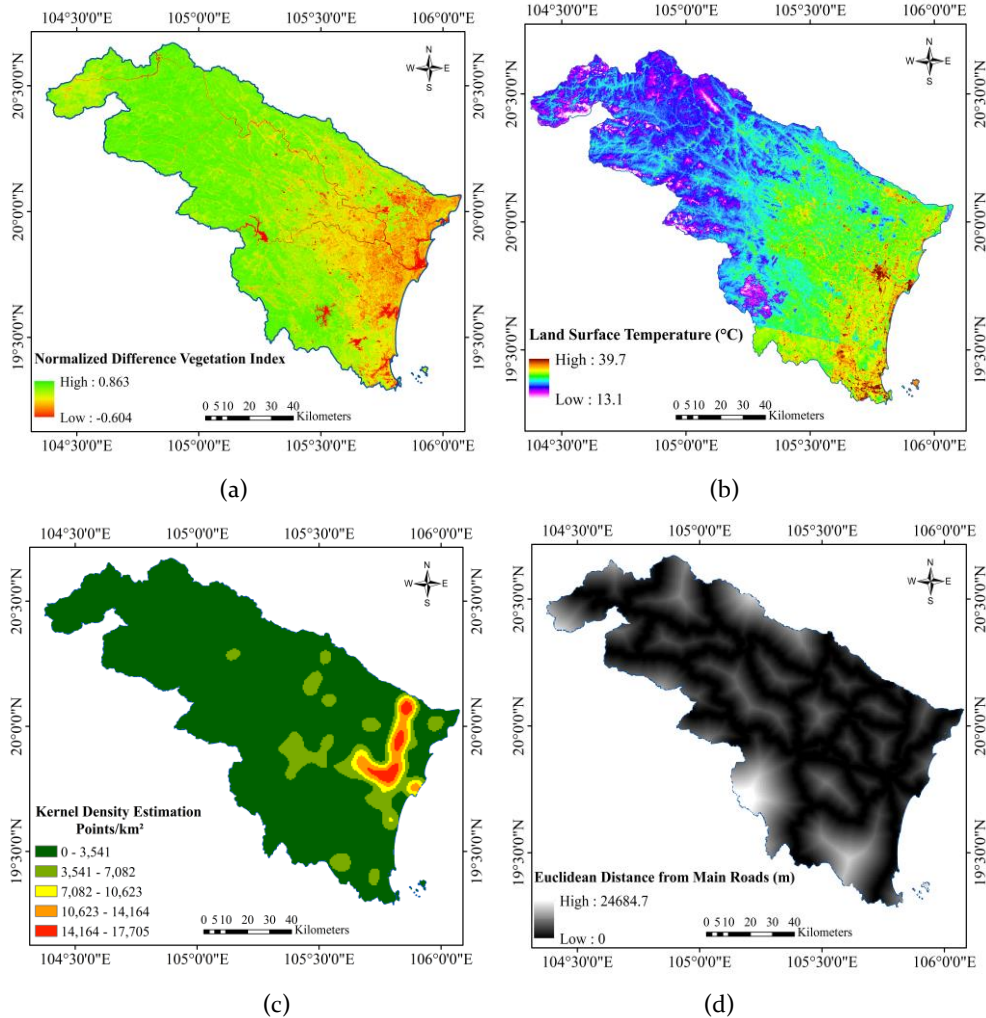


Figure 2 Input data layers for the accident risk zoning model (a) Normalized Difference Vegetation Index (b) Land Surface Temperature (c) Kernel Density Estimation (d) Euclidean Distance from Main Roads

METHODOLOGY

This study employs two parallel approaches to analyze traffic accident risk: the WOA and RF machine learning model. Each method utilizes distinct strategies for spatial data processing, input variable selection, and risk map generation. The overall analytical workflow, illustrated in Figure 3, comprises three main stages:

(i) Data Collection and Preprocessing: This stage involves the acquisition and preparation of spatial data layers, including those derived from remote sensing imagery and base maps.

(ii) Independent Risk Analyses: The WOA method normalizes and assigns weights to raster layers based on expert judgment to produce a composite risk map. Concurrently, the RF model is trained using actual accident

point data along with synthetically generated non-accident points, allowing for classification and performance evaluation.

(iii) Comparative Assessment: The final stage involves a comparative evaluation of the two approaches to determine their predictive accuracy and to identify the spatial factors most strongly associated with traffic accident risk.

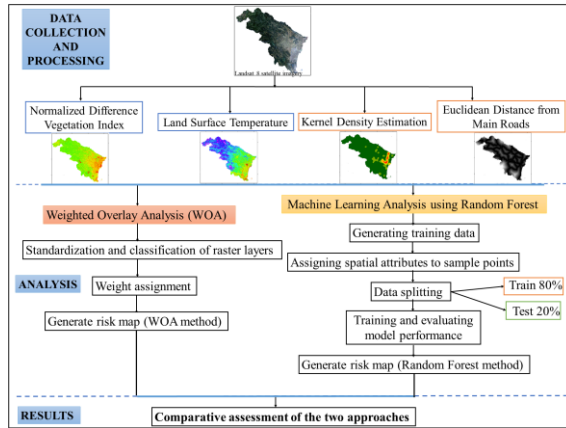


Figure 3. Analysis flowchart of WOA and RF

All data processing, spatial analysis, and model construction are conducted within the ArcGIS Pro software environment. Key tools employed include Weighted Overlay for assigning weights to spatial input layers, KDE for calculating accident density, and the Train Random Trees Classifier for implementing the RF model to generate risk zoning maps.

Weighted overlay analysis

WOA is a widely used technique in GIS for solving multi-criteria spatial decision-making problems. It is particularly effective in evaluating the relative suitability of locations based on multiple heterogeneous input factors. In this study, WOA is applied to delineate areas with elevated traffic accident risk.

Each input layer is reclassified on a standardized scale from 1 to 5, representing the relative influence or risk level of each spatial factor. These reclassified raster layers are then normalized and combined using a weighted sum approach, where each cell value is computed based on its contribution to the

overall risk, as determined by predefined expert-assigned weights. The final composite risk map is generated using the following weighted sum equation (Equation 1)(Doke et al., 2021):

$$S = \sum_{i=1}^n w_i \cdot x_i \quad (1)$$

where:

S : composite risk index;

x_i : standardized value of the i -th layer;

w_i : weight of the i -th layer;

n : number of data layers.

WOA offers several advantages: it is intuitive, straightforward to implement within the ArcGIS environment, and supports flexible integration of heterogeneous spatial data sources-including remote sensing imagery, topographic maps, GPS data, and field surveys. However, a key limitation of the method lies in its reliance on accurate weight assignment for each input layer. When empirical data are unavailable for automatic optimization, weights must be assigned manually based on expert judgment, domain knowledge, previous studies, or regional characteristics (Malczewski et al., 2006).

In this study, after standardizing all input layers, they are integrated using the WOA technique. The specific weights assigned to each layer are detailed in Table 1.

Table 1. Weights of criteria classes in WOA analysis (Leake et al., 2000)

Criteria	Weight (%)
Kernel Density Estimation	30%
Land Surface Temperature	20%
Normalized Difference Vegetation Index	15%
Euclidean Distance from Main Roads	35%

Segmentation using the Random Forest Model

RF is a robust ensemble learning algorithm that constructs multiple decision

trees to improve classification accuracy and mitigate overfitting (Breiman, 2001). Each tree is trained on a bootstrap sample of the dataset, and at each decision node, a random subset of predictor variables is considered for splitting. This mechanism promotes model diversity and enhances overall stability and generalizability.

In this study, the RF model was trained on a dataset comprising 3,188 labeled points—consisting of 1,594 actual traffic accident locations and 1,594 randomly sampled non-accident locations. Feature values for these points were extracted from four spatial raster layers: NDVI, LST, KDE, and Euclidean Distance from Main Roads. This was accomplished using the Extract Multi Values to Points tool within ArcGIS Pro. The dataset was randomly split into training (80%) and testing (20%) subsets, ensuring balanced class representation.

Model training was performed directly in ArcGIS Pro using the Train Random Trees Classifier tool, enabling tight integration between spatial data processing and machine learning implementation. The RF model not only demonstrated strong performance in classifying accident-prone areas but also provided variable importance scores, which offer valuable insights for evidence-based traffic planning and risk mitigation strategies (Rodriguez-Galiano et al., 2012; Obasi et al., 2023).

Model evaluation metrics

In traffic accident risk classification, datasets often exhibit class imbalance, which can distort evaluation if only overall accuracy is considered. To provide a more nuanced assessment, additional performance metrics such as Precision, Recall, and F1-score are employed. These metrics are particularly effective for evaluating classification performance under imbalanced conditions, ensuring that both false positives and false negatives are appropriately accounted for (Yan et al., 2022).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

In which:

TP (True Positive): correctly predicted a positive sample;

TN (True Negative): correctly predicted a negative sample;

FP (False Positive): incorrectly predicted a negative sample as positive;

FN (False Negative): incorrectly predicted a positive sample as negative.

RESULTS AND DISCUSSION

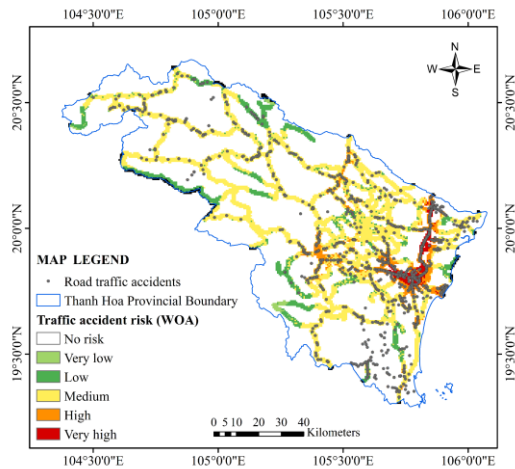
Accident risk zoning using WOA and Random Forest

Both the WOA and RF models were applied to classify traffic accident risk zones across Thanh Hoa Province. The WOA-derived risk map as shown in Figure 4(a) indicates that high-risk zones are primarily concentrated along major transportation corridors and in densely populated urban centers. This result aligns with intuitive expectations based on traffic volume and infrastructure conditions.

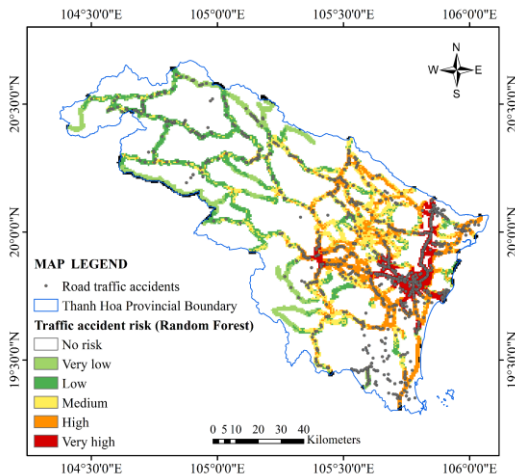
In contrast, the RF-based risk map as shown in Figure 4(b) exhibits a higher level of spatial specificity. Areas identified as high and very high risk by the RF model show strong spatial correspondence with actual accident locations. This suggests that the RF model captures the underlying spatial patterns of traffic accidents more accurately, likely due to its data-driven nature. However, in certain sparsely populated or topographically complex regions, the model may fail to detect all accident-prone locations—possibly due to the limitations of the input features or insufficient training data in those areas.

Overall, both WOA and RF methods demonstrate utility in spatial risk segmentation. The WOA method offers a rule-based, interpretable framework for decision-

making, while the RF model provides a more adaptive and empirically grounded alternative. The resulting risk maps can serve as critical tools for developing traffic warning systems, informing infrastructure investment priorities, and enhancing evidence-based traffic safety planning.



(a)



(b)

Figure 4. Traffic accident risk zoning map (a) WOA (b) RF

Evaluation of Random Forest Model Performance

The performance of the RF model was assessed using a confusion matrix, which served as the basis for calculating key classification metrics, including Accuracy, Precision, Recall, and F1-score. These metrics

provide a comprehensive evaluation of the model's predictive capability, covering aspects of correctness, sensitivity, and overall classification effectiveness on the testing dataset.

The RF model achieved an overall accuracy of 81.07%, indicating a high proportion of correctly classified instances. Precision was recorded at 87.22%, suggesting a low false positive rate, while Recall reached 72.73%, reflecting the model's ability to correctly identify accident-prone locations. The F1-score, which balances Precision and Recall, was 79.31%, confirming the model's overall robustness and reliability as shown in Table 2.

These results demonstrate that the Random Forest model delivers stable and effective performance, making it a strong candidate for real-world applications in traffic accident risk prediction, spatial planning, and the development of early-warning systems.

Table 2 Percentage of evaluation indicators of Random Forest model

Indicator	Percentage
Accuracy	81,07%
Precision	87,22%
Recall	72,73%

role of spatial factors

In the WOA, input layer weights were assigned based on expert judgment as shown in Table 1, reflecting domain knowledge about the relative influence of each factor on traffic accident risk. While effective, this approach inherently introduces subjectivity. To provide a more objective and data-driven perspective, the study also utilized the Train Random Trees Model tool in ArcGIS Pro to quantify the relative importance of each variable within the RF model as shown in Table 3.

The RF results broadly align with the expert-assigned WOA weights, particularly regarding the two most influential spatial factors: Euclidean Distance from Main Roads and KDE, as illustrated in Figure 6. Among all

variables, KDE was the most important factor, contributing 38.33%—significantly higher than the 30% in WOA. This reinforces the established understanding that spatial clustering of past accidents is a strong predictor of future risk, as validated in recent KDE-based spatial analyses (Le et al., 2024a; Le et al., 2024b).

The second most important factor was the Euclidean distance to main roads, accounting for 27.67% of the model's influence, lower than the 35% in WOA but still underscoring the critical role of proximity to main roads—typically with higher traffic volumes and speeds—in increasing accident risk. This result is consistent with prior studies, which demonstrate that proximity to primary roadways strongly correlates with accident frequency and severity due to higher traffic volumes and greater exposure to high-speed travel (Wang et al., 2023; Wang et al., 2019). Together, these two features alone explain nearly two-thirds of the model's predictive capability, highlighting their dominant role in risk stratification.

Environmental and land cover factors also contributed meaningfully: LST and NDVI contributed 19.50% and 14.50%, respectively. These variables influence accident risk through clear physical mechanisms. Elevated LST can reduce pavement friction and contribute to driver fatigue, while variations in NDVI affect sight distance, shading, and the presence of roadside vegetation—conditions known to alter driving behavior and increase the likelihood of traffic accidents. While secondary to the top two features, these variables still offer valuable context - especially in urban settings, where LST reflects heat island effects that may influence driver behavior and vehicle performance (Pińskwar et al., 2024). NDVI, as a proxy for vegetation cover and landscape type, also indirectly captures aspects of road visibility, surface conditions, and land use patterns.

The comparison between WOA-assigned weights and RF-based importance values highlights the complementary strengths of the

two approaches. Expert weighting captures nuanced domain insights, while machine learning quantifies actual data-driven influence. Therefore, combining both perspectives - expert-driven and model-derived - enhances the overall reliability, objectivity, and interpretability of the traffic accident risk zoning process.

Table 3. Variable importance of input factors in the Random Forest model

Input Variable	Importance (%)
Kernel Density Estimation	38.33%
Euclidean Distance from Main Roads	27.67%
Land Surface Temperature	19.50%
Normalized Difference Vegetation Index	14.50%

Comparison between Weighted Overlay Analysis and Random Forest

To assess the effectiveness of the two methods -WOA and RF - accident point data were spatially overlaid with each method's risk zoning map as shown in Figure 4(a) and Figure 4(b). The WOA and RF models operate on fundamentally different scales: WOA uses a standardized 1–5 index, whereas RF produces probability-based outputs ranging from 0 to 1. Therefore, a unified classification threshold cannot be applied. Each model was categorized according to its intrinsic value distribution. However, the comparison was conducted using harmonized indicators—such as the percentage of accident points captured within high-risk zones and the RF performance metrics—to ensure an objective and scale-independent evaluation. Additionally, the “no risk” class assigned to mountainous and forested areas without traffic networks is applied as a spatial mask rather than a model-generated category. Therefore, this adjustment does not modify the underlying WOA or RF scales or affect the harmonized comparative evaluation presented above.

The comparative analysis reveals that the RF model provides a more accurate classification of high-risk areas. Specifically, 80.61% of actual accident locations fall within

the two highest risk categories as defined by the RF model. In contrast, the WOA method captures only 71.45% of accident points within the same categories.

Beyond spatial accuracy, the RF model also offers quantitative evaluation metrics, including an overall accuracy of 81.07%, along with precision, recall, and F1-score, which enable a more rigorous validation of its predictive performance. Moreover, the RF model produces variable importance scores, offering interpretable insights into the contribution of each input factor—an advantage that WOA, being a rule-based method, lacks.

These findings underscore the superiority of the machine learning approach in both predictive capability and diagnostic transparency. Unlike WOA, which is based entirely on expert-assigned weights and therefore more vulnerable to subjective bias, the RF model is data-driven and adaptive, capable of capturing complex, non-linear relationships among variables.

That said, WOA retains value in specific contexts. It is straightforward to implement, interpretable, and especially useful in scenarios where data are scarce or machine learning tools are not available. When used in combination with RF, WOA can serve as a complementary method—either to validate RF outputs or to provide initial estimates in data-constrained environments.

A summary comparison of the two approaches, based on key performance and methodological criteria, is presented in Table 4.

Table 4. Comparison of the performance of the two methods WOA and RF

Criteria	WOA	Random Forest
Requires real accident data	Yes	No
Classification performance	Medium	High
Explainability	Yes (weighting)	Explicit (determining the influence of the variable)
Degree of automation	Low	High

CONCLUSION

This study demonstrates that integrating GIS and remote sensing data with both WOA and the RF machine learning model enables a robust and comprehensive approach to traffic accident risk assessment. Among the two methods, the RF model outperforms WOA in terms of classification accuracy and variable interpretability, confirming its suitability for data-driven spatial analysis. Notably, KDE emerged as the most influential predictor of accident risk, reinforcing its importance in transportation planning. While the WOA method remains a practical choice in contexts with limited field data - due to its simplicity, transparency, and rapid deployment - it lacks the objectivity and predictive power of machine learning models. Nevertheless, its use as a complementary tool can enhance initial assessments or serve as a benchmark in resource-constrained settings.

The findings strongly support the integration of machine learning algorithms with GIS and remote sensing technologies in future transportation safety initiatives. This combined framework can significantly improve spatial risk prediction, support proactive infrastructure planning, and enable the development of early-warning systems to mitigate road traffic accidents. Future research should expand this approach by incorporating additional spatial variables such as road geometry, traffic volume, and population density, and by exploring advanced ensemble or deep learning models to further enhance analytical precision and temporal scalability.

REFERENCES

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1). <https://doi.org/10.1023/A:1010933404324>.
- Chen, H., Wood, M. D., Linstead, C., & Maltby, E. (2011). Uncertainty analysis in a GIS-based multi-criteria analysis tool for river catchment management. *Environmental Modelling and Software*, 26(4). <https://doi.org/10.1016/j.envsoft.2010.09.005>.
- Doke, A. B., Zolekar, R. B., Patel, H., & Das, S. (2021). Geospatial mapping of groundwater potential zones using multi-criteria decision-making AHP approach in a

- hardrock basaltic terrain in India. *Ecological Indicators*, 127. <https://doi.org/10.1016/j.ecolind.2021.107685>.
- Eskandari, H., & Rabelo, L. (2007). Handling uncertainty in the analytic hierarchy process: A stochastic approach. *International Journal of Information Technology and Decision Making*, 6(1). <https://doi.org/10.1142/S0219622007002356>.
- Farhangi, F., Sadeghi-Niaraki, A., Razavi-Termeh, S. V., & Choi, S. M. (2021). Evaluation of tree-based machine learning algorithms for accident risk mapping caused by driver lack of alertness at a national scale. *Sustainability (Switzerland)*, 13(18). <https://doi.org/10.3390/su131810239>.
- Le, K. G., Tran, Q. H., & Do, V. M. (2024). Urban Traffic Accident Features Investigation to Improve Urban Transportation Infrastructure Sustainability by Integrating GIS and Data Mining Techniques. *Sustainability (Switzerland)*, 16(1). <https://doi.org/10.3390/su16010107>.
- Le, T. H., Vu, T. P. T., & Do, T. P. T. (2024a). Analyzing Road Traffic Incident Hotspots Using Cluster Analysis in Thanh Hoa Province of Vietnam. *International Journal of Geoinformatics*, 20(9), 16–26. <https://doi.org/10.52939/ijg.v20i9.3537>.
- Le, T. H., Vu, T. P. T., & Do, T. P. T. (2024b). A study on identifying and analyzing road traffic incident hotspots on National Highway 1A, Thanh Hoa province, Vietnam, employing Statistical and GIS Techniques. *Journal of Mining and Earth Sciences*, 65(6), 22–33. [https://doi.org/10.46326/JMES.2024.65\(6\).03](https://doi.org/10.46326/JMES.2024.65(6).03).
- Leake, C., & Malczewski, J. (2000). GIS and Multicriteria Decision Analysis. *The Journal of the Operational Research Society*, 51(2). <https://doi.org/10.2307/254268>.
- Liu, L., Gao, Z., Luo, P., Duan, W., Hu, M., Mohd Arif Zainol, M. R. R., & Zawawi, M. H. (2023). The Influence of Visual Landscapes on Road Traffic Safety: An Assessment Using Remote Sensing and Deep Learning. *Remote Sensing*, 15(18). <https://doi.org/10.3390/rs15184437>.
- Malczewski, J. (2006). GIS-based multicriteria decision analysis: A survey of the literature. *International Journal of Geographical Information Science*, 20(7). <https://doi.org/10.1080/13658810600661508>.
- Obasi, I. C., & Benson, C. (2023). Evaluating the effectiveness of machine learning techniques in forecasting the severity of traffic accidents. *Heliyon*, 9(8). <https://doi.org/10.1016/j.heliyon.2023.e18812>.
- Pińskwar I, Choryński A, & Graczyk D 2024 *Good weather for a ride (or not?): how weather conditions impact road accidents—a case study from Wielkopolska (Poland)*, *International journal of biometeorology*, 68(2), 317–331. <https://doi.org/10.1007/s00484-023-02592-3>.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(1). <https://doi.org/10.1016/j.isprsjprs.2011.11.002>.
- Vu. T. P. T., Dang. T. G., & Le. V. A., (2024). Reliability assessment of land subsidence monitoring results using PSI technique in Ho Chi Minh City, Vietnam. *International Journal of Environmental Studies*, 81(2), 881–895. <https://doi.org/10.1080/00207233.2024.2324623>.
- Wang, J., Ma, S., Jiao, P., Ji, L., Sun, X., & Lu, H. (2023). Analyzing the Risk Factors of Traffic Accident Severity Using a Combination of Random Forest and Association Rules. *Applied Sciences (Switzerland)*, 13(14). <https://doi.org/10.3390/app13148559>.
- Wang, W., Yuan, Z., Yang, Y., Yang, X., & Liu, Y. (2019). Factors influencing traffic accident frequencies on urban roads: A spatial panel time-fixed effects error model. *PLOS ONE*, 14(4). <https://doi.org/10.1371/journal.pone.0214539>.
- Worachairungreung, M., Ninsawat, S., Witayangkurn, A., & Dailey, M. N. (2021). Identification of road traffic injury risk prone area using environmental factors by machine learning classification in Nonthaburi, Thailand. *Sustainability (Switzerland)*, 13(7). <https://doi.org/10.3390/su13073907>.
- Yan, M., & Shen, Y. (2022). Traffic Accident Severity Prediction Based on Random Forest. *Sustainability (Switzerland)*, 14(3). <https://doi.org/10.3390/su14031729>.
- Zheng, M., Xie, X., Jiang, Y., Shen, Q., Geng, X., Zhao, L., & Jia, F. (2024). Optimizing Kernel Density Estimation Bandwidth for Road Traffic Accident Hazard Identification: A Case Study of the City of London. *Sustainability*, 16(16), 6969. <https://doi.org/10.3390/su16166969>.