TẠP CHÍ KHOA HỌC VÀ CÔNG NGHỆ ĐẠI HỌC DUY TÂN
DTU Journal of Science and Technology 01(68) (2025) 26-35

ĐẠI HỌC DUY TÂN
DUYTAN UNIVERSITY

Machine learning-based geospatial analysis for pluvial flood susceptibility mapping in Da Nang urban area

Phân tích không gian địa lý dựa trên học máy cho việc lập bản đồ phân vùng nhạy cảm ngập lụt do mưa ở khu vực đô thị Đà Nẵng

Hoang Nhat Duc[a,b*], Nguyen Quoc Lam[b], Pham Quang Nhat[b]
Hoàng Nhật Đức[a,b*], Nguyễn Quốc Lâm[b], Phạm Quang Nhật[b]

[a]*Institute of Research and Development, Duy Tan University, Da Nang, 550000, Vietnam*
[a]*Viện Nghiên cứu và Phát triển Công nghệ cao, Đại học Duy Tân, Đà Nẵng, Việt Nam*
[b]*Faculty of Civil Engineering, School of Engineering and Technology, Duy Tan University, Da Nang, 550000, Vietnam*
[b]*Khoa Xây dựng, Trường Công nghệ và Kỹ thuật, Đại học Duy Tân, Đà Nẵng, Việt Nam*

(Date of receiving article: 30/07/2024, date of completion of review: 02/10/2024, date of acceptance for posting: 11/11/2024)

Abstract

Urban flooding in Da Nang has become increasingly frequent, resulting in significant damages to both human life and property. Despite the growing concern, assessments of flood risk in this region remain limited. This study aims to construct a pluvial flood susceptibility map for the Da Nang urban area by utilizing flood locations reported in reliable and official sources of local mass media to create a flood inventory. The current work collects remote sensing data and performs geospatial analysis to develop thematic maps that represent various influencing factors of flood susceptibility, including built-up density, distance to rivers, distance to retention ponds, distance to drainage canals, elevation, topographic position index, slope, normalized difference vegetation index, bare soil index, and normalized difference built-up index. The extreme gradient boosting machine (XGBoost) model is employed to assess flood susceptibility in the study area, achieving a classification accuracy rate of approximately 93%. The resulting pluvial flood susceptibility map serves as a valuable tool for local authorities, enabling them to implement effective flood mitigation strategies, enhance urban planning, and prioritize infrastructure improvements to reduce flood risks and protect vulnerable communities.

Keywords: Urban flooding; pluvial flood; geospatial analysis; machine learning; remote sensing.

Tóm tắt

Lũ lụt đô thị ở Đà Nẵng đã trở nên ngày càng thường xuyên, gây ra những thiệt hại đáng kể về cả con người và tài sản. Dù vậy, việc đánh giá rủi ro lũ lụt ở khu vực này vẫn còn hạn chế. Mục tiêu của bài báo này là xây dựng bản đồ nguy cơ lũ lụt do mưa cho khu vực đô thị Đà Nẵng bằng cách sử dụng các vị trí lũ lụt được báo cáo trong các nguồn tin chính thống và đáng tin cậy của phương tiện truyền thông địa phương. Các thông tin thu thập được dùng để xây dựng một cơ sở các vị trí lũ lụt trong đô thị. Nghiên cứu của chúng tôi cũng thu thập dữ liệu viễn thám và thực hiện phân tích không gian địa lý để phát triển các bản đồ chuyên đề đại diện cho các yếu tố ảnh hưởng đến nguy cơ lũ lụt. Các yếu tố này bao gồm mật độ xây dựng, khoảng cách đến sông, khoảng cách đến các ao giữ nước, khoảng cách đến các kênh thoát nước, độ cao, chỉ số vị trí địa hình, độ dốc, chỉ số thảm thực vật, chỉ số đất trống, và chỉ số xây dựng. Mô hình học máy XGBoost được sử dụng để đánh giá nguy cơ lũ lụt trong khu vực nghiên cứu, giúp đạt được độ chính xác trong phân loại khoảng

Corresponding author: Hoang Nhat Duc
Email: hoangnhatduc@duytan.edu.vn

93%. Do đó, bản đồ nguy cơ lũ lụt được xây dựng trong nghiên cứu này là một công cụ tham khảo hữu ích cho các cơ quan địa phương, giúp họ thực hiện các chiến lược giảm thiểu tác hại của lũ lụt một cách hiệu quả, cải thiện quy hoạch đô thị, cải tiến cơ sở hạ tầng nhằm giảm thiểu rủi ro lũ lụt, và bảo vệ các cộng đồng tại những nơi có nguy cơ lụt lội cao.

Từ khóa: Lụt lội đô thị; lụt do mưa lớn; phân tích không gian địa lý; học máy; viễn thám.

1. Introduction

Urban flooding has emerged as a significant challenge for cities worldwide, characterized by the inundation of land and property due to excessive rainfall overwhelming drainage systems [9]. This phenomenon typically occurs in a city located near rivers or coastlines, with various types of flooding (e.g., pluvial, fluvial, and coastal flooding). Each type of flood presents unique causes and impacts. The increasing frequency of urban flooding events is closely linked to rapid urbanization and climate change, which increases the vulnerabilities of urban infrastructure and populations. As cities expand and the area of impervious surfaces like roads and buildings increases, the natural ability of the landscape to absorb rainfall diminishes, leading to heightened surface runoff and flooding risks.

The consequences of urban flooding are profound, affecting human safety, property, and economic stability [5]. Flood events can result in loss of life, displacement of communities, and extensive damage to infrastructure and homes. Furthermore, the economic repercussions can be severe, with costs associated with emergency response, recovery, and rebuilding efforts. Climate change is anticipated to intensify these challenges, as abnormal rainfall patterns and rising sea levels contribute to more frequent and severe flooding events. As such, urban flooding not only poses immediate threats but also raises long-term concerns regarding urban resilience and sustainability, necessitating proactive measures for flood risk assessment and management in urban planning.

In recent years, geospatial analysis, GIS technology, and remote sensing data increasingly play a pivotal role in assessing flood risk and susceptibility in urban areas [4]. By integrating various data sources, these technologies enable researchers and urban planners to analyze complex environmental factors that contribute to flooding [7]. GIS allows for the visualization and analysis of spatial data, facilitating the identification of flood-prone areas based on variables such as land use, topography, and hydrological conditions [3]. Remote sensing data, obtained from satellites and aerial imagery, provides critical information on rainfall patterns, surface conditions, and changes in land cover, enhancing the understanding of how these factors influence flood dynamics.

Additionally, the combination of geospatial analysis and machine learning techniques further enhances flood susceptibility assessments by enabling the modeling of intricate relationships between various influencing factors [3]. This approach allows for the development of detailed flood susceptibility maps that can inform decision-making processes for urban planning and disaster management. By identifying areas at high risk of flooding, local authorities can implement targeted mitigation strategies, optimize infrastructure investments, and enhance community resilience against future flood events. Overall, the application of geospatial analysis, GIS, and remote sensing is essential for developing effective flood risk management frameworks in increasingly vulnerable urban environments.

Da Nang is a crucial city located in the coastal area of central region in Vietnam. This city is increasingly vulnerable to flood hazards, particularly urban flooding, which poses significant challenges to its infrastructure and residents. The city has experienced a notable rise in flooding incidents, caused by rapid urbanization and climate change. Annual floods bring about severe impacts on both human safety and property during heavy rainfall. The combination of an aging drainage system, increased impervious surfaces due to urban development, and the accumulation of waste blocking drainage channels contributes to the city's susceptibility to flooding.

According to a report from the local government, flooding in Da Nang City between 1998 and 2020 resulted in significant human and economic losses [8]. During this period, flooding claimed the lives of 83 people and injured 100 others. Additionally, 15,633 households were impacted, and 130,861 houses were destroyed, leading to an economic loss exceeding 2,600 billion VND [8]. The report highlights that flooding has been the most devastating natural disaster in the history of Da Nang City. As a result, local authorities are under pressure to implement effective flood management strategies to mitigate risks and enhance the resilience of the urban environment against future flood events.

Based on such motivations, this research aims to establish a pluvial flood susceptibility map for the Da Nang urban region by utilizing flood data sourced from reliable sources of local media to compile a flood inventory. The study gathers remote sensing information and conducts spatial analysis to generate thematic maps that illustrate various factors influencing flood susceptibility, such as built-up density, proximity to rivers, retention ponds, drainage canals, elevation, topographic position index, slope, normalized difference vegetation index, bare soil index, and normalized difference built-up index. The extreme gradient boosting machine (XGBoost), a state-of-the-art machine learning model, is then applied to evaluate flood susceptibility across the study area.

2. Research method and materials

2.1. Study area

Da Nang is located in the regions in Central Vietnam and is situated downstream of the Vu Gia - Thu Bon River system [1]. This area has faced numerous severe flooding events and is frequently inundated during the rainy season [6]. Pluvial and fluvial flooding events have become an annual occurrence in Da Nang City, especially under the influence of recent climate change, leading to increasingly severe flood damage. The study area of the paper encompasses the inland districts of Da Nang City (see Figure 1).

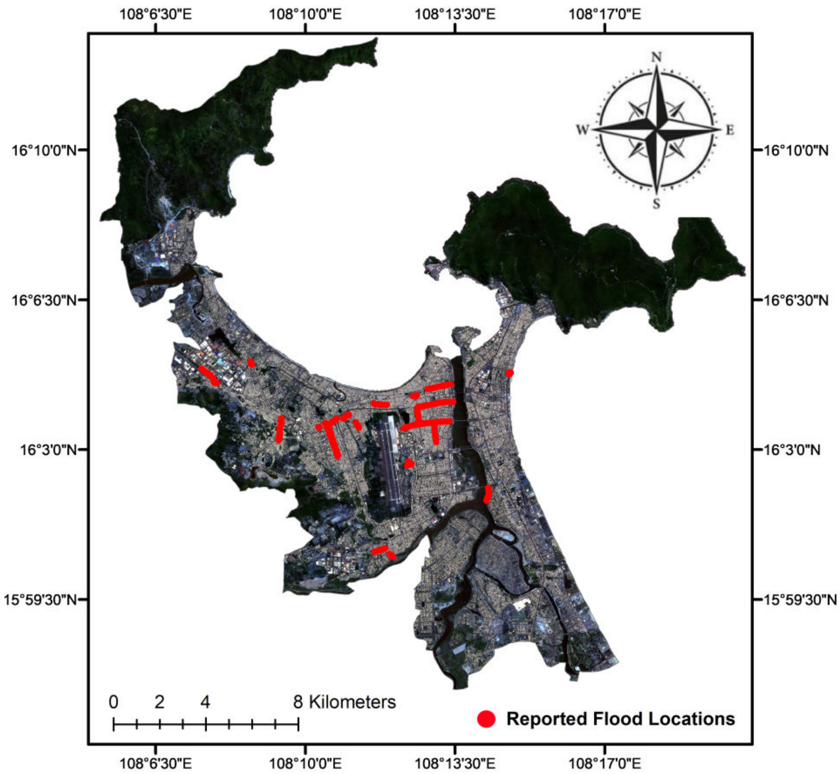

Figure 1. The study area of Da Nang's urban center

The Hoang Sa Islands and other rural districts of Hoa Vang are excluded from this research as they are not part of the urban center being examined. This region is characterized by two primary river systems, the Cu De and Han Rivers. It experiences two distinct seasons: a rainy season from August to December and a dry season from January to July. Annually, the area is susceptible to typhoons and floods, particularly from September to December [8]. The topography of the study area is diverse, with elevations ranging from -26 to 1,170 meters above sea level, extending from mountainous regions in the west to low-lying areas in the east. The relatively short distance between the mountains to the west and the coastline to the east contributes to the heightened flood risk in this region.

Moreover, constructing a flood inventory is a crucial first step in developing an accurate flood susceptibility map using machine learning techniques, GIS, geospatial analysis, and remote sensing data. The flood inventory serves as the foundation for training and validating the predictive models, ensuring that the resulting susceptibility map reflects the actual flood occurrences in the study area. By collecting data on past flood events from reliable sources, including local mass media reports, the flood inventory captures the spatial distribution and characteristics of flooding in the region. This information is then integrated with geospatial datasets representing various influencing factors. Machine learning can then analyze the complex relationships between these factors and the flood inventory to identify patterns and predict areas susceptible to future flooding. This study has reviewed reliable source of local mass media reports to identify confirmed flood locations occurred in 2022 and 2023. The detailed reports of flood locations are provided in Table 1.

Table 1. Reported flood locations

Flood locations	Periods	Sources
Hoang Dieu, Nguyen Van Linh, Hung Vuong, Quang Trung, Hue intersection (Nga Ba Hue-Ton Duc Thang), Me Suot	October of 2022	https://www.qdnd.vn/xa-hoi/cac-van-de/da-nang-quyet-liet-cac-giai-phap-chong-ngap-lut-753272
K634, K640, K654 Trung Nu Vuong	October of 2023	https://tuoitre.vn/dan-vung-ngap-da-nang-nong-long-thoat-lut-20231018092221482.htm
Nguyen Nhan	October of 2022	https://tuoitre.vn/da-nang-khang-dinh-co-giai-phap-can-co-giai-quyet-ngap-lut-do-thi-20231025214810138.htm
Truong Son, Chuong Duong, Nguyen Hoang, Nguyen Van Linh, Nguyen Tri Phuong	September of 2023	https://congan.danang.gov.vn/-/-a-nang-truoc-nhung-thach-thuc-an-ninh-phi-truyen-thong
Ton Duc Thang, Tan Trao, 140 Quang Trung, Ton Duc Thang, Bac Son, 12 Nam Tran, Vuong Thua Vu-Vo Nguyen Giap, 102/7 Dien Bien Phu, 15 Phan Boi Chau, 17 Tran Phu, Ham Nghi, Hai Phong (Da Nang railway station)	October of 2023	https://nld.com.vn/thoi-su/da-nang-ngap-sau-luong-mua-gan-dat-dinh-ngay-mua-lich-su-20231014075458078.htm
Lac Long Quan, Au Co, Dong Ke, Road No. 4, Nguyen Luong Bang (Gate of Hoa Khanh industrial zone), Thanh Vinh market (Au Co)	November of 2023	https://nhandan.vn/anh-nhieu-tuyen-duong-tai-da-nang-ngap-sau-sau-mua-lon-post781443.html

2.2. Influencing factors

The assessment of pluvial flood susceptibility in urban areas is critical for effective disaster risk management and urban planning. Various influencing variables play a significant role in determining the vulnerability of specific locations to flooding, particularly in the context of increasing urbanization and climate change. Among these variables, built-up density, proximity to water bodies (such as rivers, retention ponds, and drainage canals), elevation, topographic position index, slope, and vegetation indices (like the normalized difference vegetation index and bare soil index) are selected in the current study for creating accurate susceptibility maps. These factors were selected based on the review of existing works [4,8,7] as well as the availability of data in the study area.

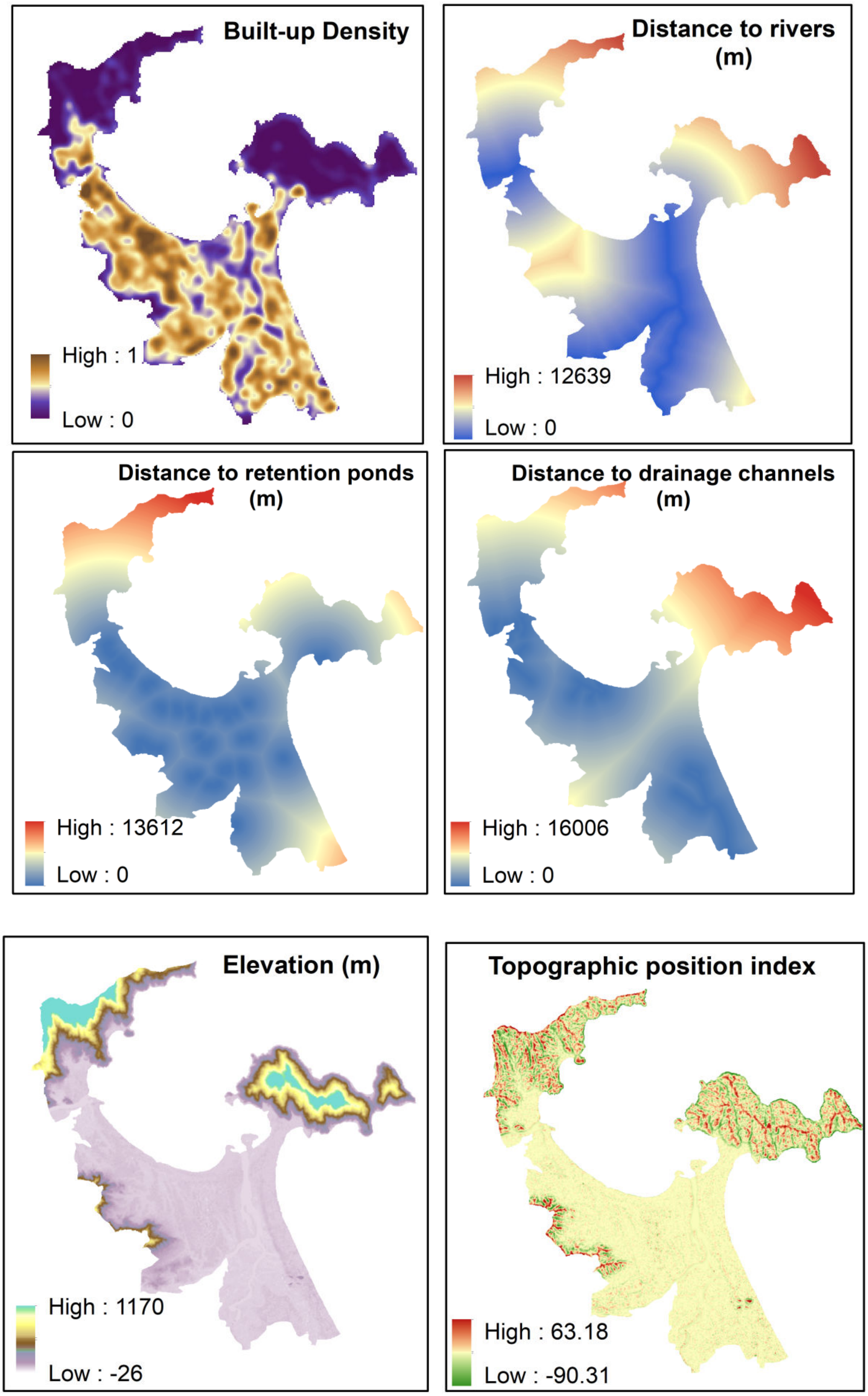

Figure 2. Thematic maps of the study area

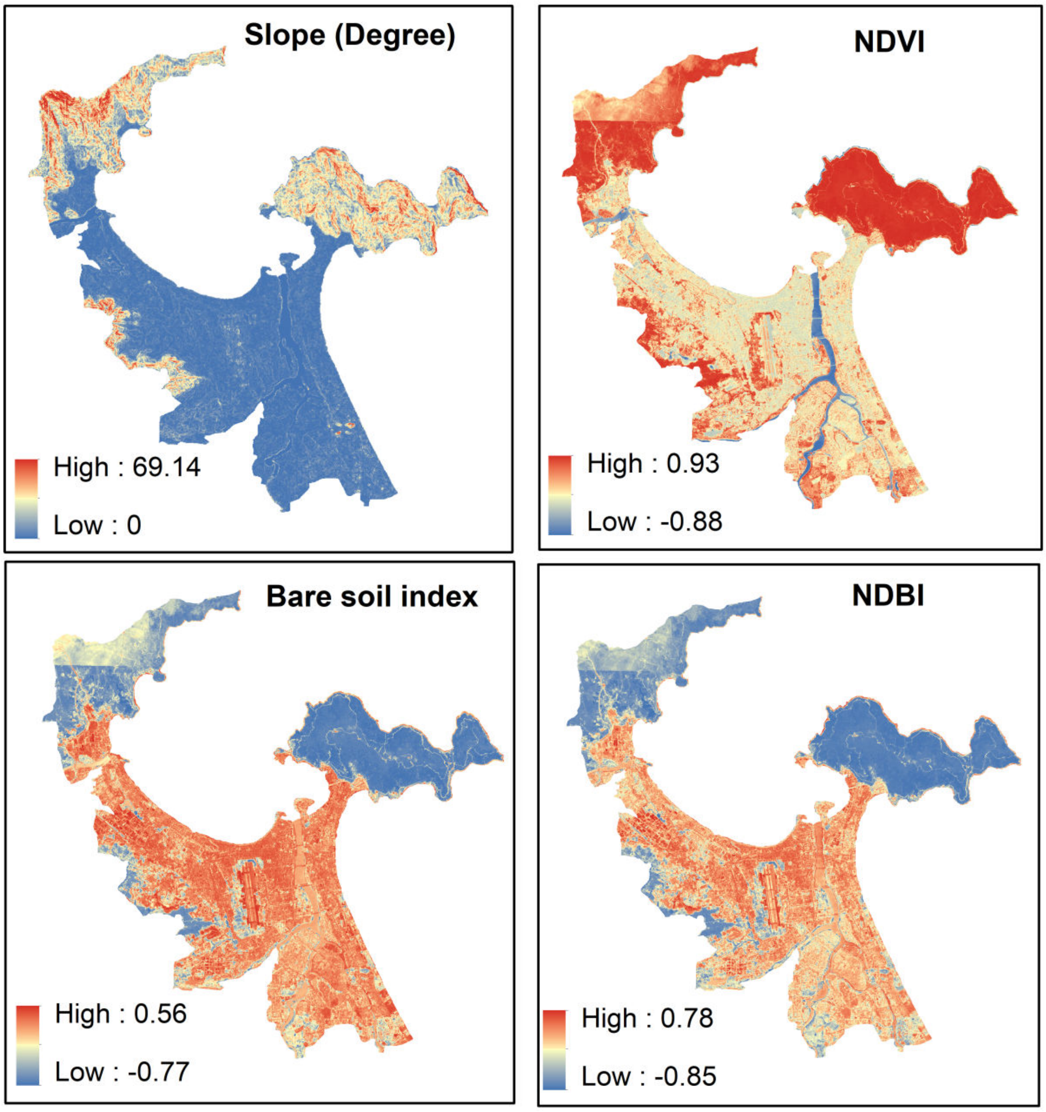

Figure 2. (cont.)

The influencing variables are crucial; the machine learning model can learn the patterns of those variables for identifying areas at higher risk of flooding. Built-up density indicates how much impervious surface exists, which affects runoff and drainage. Proximity to rivers and drainage systems can determine how quickly water can be evacuated from an area during heavy rainfall. Elevation and slope influence water flow and accumulation, while the topographic position index helps to classify landforms and their potential for water retention or runoff. Vegetation indices provide insights into land cover and soil moisture, which are vital for understanding how much rainfall can be absorbed versus how much will contribute to surface runoff. Bare soil index can be used to identify areas with little to no vegetation cover. The areas having high values of the bare soil index are more susceptible to runoff and erosion. Meanwhile, the normalized difference built-up index (NDBI) helps differentiate built-up areas from natural landscapes, providing insight into urbanization's impact on flood susceptibility. The thematic maps of the study area are provided in Figure 2. The normalized difference vegetation index (NDVI), NDBI, and bare soil index are computed from Sentinel 2's bands and processed in Google Earth Engine's code editor. The factors of elevation, slope, and topographic position index are computed from the digital elevation data obtained from the NASA SRTM Digital Elevation dataset. The maps are prepared in ArcGIS package.

2.3. Machine learning approach

This study utilizes the XGBoost for susceptibility mapping of pluvial floods in Da Nang city. The XGBoost [2] is a powerful machine learning algorithm widely used for classification tasks due to its high performance and efficiency. It is an implementation of gradient-boosted decision trees designed to optimize speed and accuracy, making it particularly effective for structured or tabular data. XGBoost works by sequentially adding new models to correct errors made by existing ones, utilizing a technique known as boosting.

This approach allows the model to effectively minimize loss functions through gradient descent, resulting in robust predictive capabilities for pattern classification problems. For binary classification tasks, XGBoost uses the log loss function for training the model's structure. To prevent overfitting, this machine learning method incorporates L_1 and L_2 regularization techniques which improve model generalization by penalizing overly complex models. In this study, the model construction and prediction phases of the XGBoost model are constructed via the Python toolbox provided at https://xgboost.readthedocs.io/en/stable/.

3. Results and discussions

This study has selected a total of 500 samples from regions where flood occurrences have been reported to represent the positive class. To ensure a balanced dataset, an additional 500 samples were collected from areas with no recorded flood events within the study region. This resulting dataset comprises 1,000 samples and includes 10 explanatory variables, which are utilized to train and validate the XGBoost model. Furthermore, the performance of the XGBoost model for pluvial flood susceptibility mapping is quantitatively assessed using several metrics, including the classification accuracy rate, precision, recall, F1 score, and Cohen's Kappa coefficient. The dataset was randomly partitioned into a training set (90%) and testing set (10%). The model's evaluation process has been repeated 20 times to negate the effect of randomness in data sampling. The average performance of the model is reported as follows: classification accuracy rate = 0.93, precision = 0.90, recall = 0.96, F1 score = 0.93, and Cohen's Kappa coefficient = 0.85.

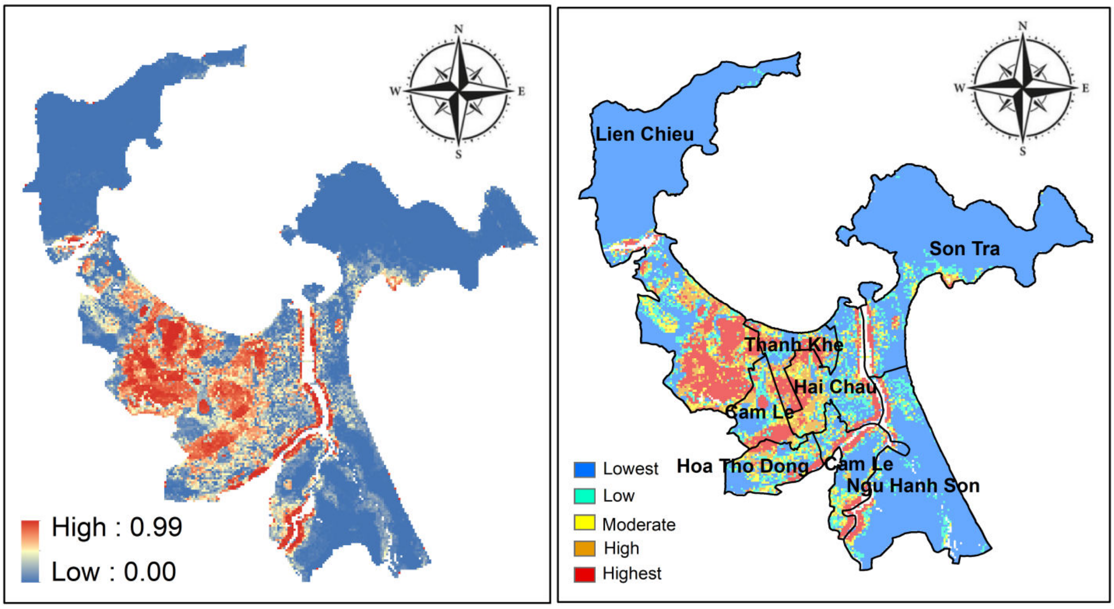

Figure 3. Results of flood susceptibility mapping

The XGBoost model used for pluvial flood susceptibility mapping in this study demonstrates excellent performance across the used evaluation metrics. The classification accuracy rate of 0.93 indicates that the model correctly classifies 93% of the samples into their respective flood and non-flood categories. The high precision value of 0.90 suggests that the model has a low false positive rate. This fact indicates that when it predicts a data sample as a flood, it is correct 90% of the time. The recall value shows that the model is able to identify 96% of the actual flood cases. This outcome shows its strong ability to diminish false negatives, which is very desirable in flood susceptibility mapping. The F1 score, which combines precision and recall, is 0.93, further confirming the model's overall effectiveness. Additionally, the Cohen's Kappa coefficient of 0.85 demonstrates a strong agreement between the model's predictions and the actual observations, accounting for the possibility of chance agreement. These outcomes clearly show that the XGBoost model is reliable and can be used for pluvial flood susceptibility mapping in the study area.

The XGBoost model is subsequently employed to create the flood susceptibility map for the study area, as illustrated in Figure 3. The probability of the positive class, which ranges from 0 to 1, is utilized to assess the vulnerability of different areas to flooding. The map is then reclassified into five distinct categories: lowest (0.0–0.2), low (0.2–0.4), moderate (0.4–0.6), high (0.6–0.8), and highest (0.8–1.0). In areas classified by the XGBoost model as having high and highest flood risk, it is crucial to implement proactive and strategic measures to mitigate potential flood impacts. First, urban planning should prioritize the development of green infrastructure, such as permeable surfaces and green roofs, to enhance water absorption and reduce runoff.

Additionally, establishing effective drainage systems and retention ponds can help manage excess water during heavy rainfall events. Public awareness campaigns should also be initiated to inform residents about flood risks and emergency preparedness, ensuring that communities are well-informed and equipped to respond in cases of torrential rains, such as the events occurring in October of 2022. Furthermore, regular monitoring and maintenance of waterways and drainage systems are essential to prevent blockages that could intensify flooding. Finally, land-use policies should restrict development in high-risk areas, promoting the preservation of natural floodplains and wetlands, which can serve as buffers against flooding.

4. Conclusions

In conclusion, this study successfully establishes a framework for spatial modeling of pluvial flood susceptibility in Da Nang city, utilizing advanced geospatial analysis, remote sensing data, and the XGBoost. By integrating a comprehensive flood inventory and various influencing factors, the model achieves an impressive classification accuracy rate of approximately 93%. The resulting pluvial flood susceptibility map not only highlights areas at risk but also serves as a critical tool for local authorities in their efforts to implement effective flood mitigation strategies. This research emphasizes the importance of employing state-of-the-art machine learning techniques in urban flood risk assessment, offering valuable insights for enhancing urban planning and prioritizing infrastructure improvements. Ultimately, the findings of this study can contribute to better preparedness and resilience against flooding in Da Nang.

In addition to XGBoost, several other machine learning approaches can effectively predict flood susceptibility at a regional scale, such as the Adaptive Neuro-Fuzzy Inference System, deep neural networks, and other advanced gradient boosting machines. Each of these techniques has unique strengths, allowing for flexible and effective flood susceptibility assessments when integrated with geospatial data. Therefore, future research should focus on integrating these diverse methodologies with enhanced geospatial datasets to improve predictive capabilities. Additionally, exploring hybrid models that combine the strengths of various algorithms could yield more accurate and reliable flood susceptibility assessments.

References

[1] An, T.T.; Raghavan, V.; Long, N.V.; Izuru, S.; Tsutsumida, N. (2021). "A GIS-based Approach for Flood Vulnerability Assessment in Hoa Vang District, Danang City, Vietnam". *IOP Conference Series: Earth and Environmental Science 652 (1)* (012003). doi:10.1088/1755-1315/652/1/012003

[2] Chen, T.; Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System". *Paper presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA*. doi: https://doi.org/10.1145/2939672.2939785

[3] Ngo, P.T.T.; Duc, H.N.; Pradhan, B.; Khanh, N.Q.; Truong, T.X.; Minh, N.Q.; Nghia, N.V.; Samui, P.; Tien Bui, D. (2018). "A Novel Hybrid Swarm Optimized Multilayer Neural Network for Spatial Prediction of Flash Floods in Tropical Areas Using Sentinel-1 SAR Imagery and Geospatial Data". *Sensors, 18 (11)*, 3704. doi: https://doi.org/10.3390/s18113704

[4] Ngoc, N.H.; Fukuda, H.; Nguyet, N.M. (2024). "Assessment of the Susceptibility of Urban Flooding Using GIS with an Analytical Hierarchy Process in Hanoi, Vietnam". *Sustainability*, *16 (10),*3934. doi: https://doi.org/10.3390/su16103934

[5] Sörensen, J.; Mobini, S. (2017). "Pluvial, urban flood mechanisms and characteristics – Assessment based on insurance claims". *Journal of Hydrology, 555*, 51-67. doi:https://doi.org/10.1016/j.jhydrol.2017.09.039

[6] Tuan, T.H.; Tran, P.; Hawley, K.; Khan, F.; Moench, M. (2015). "Quantitative cost-benefit analysis for typhoon resilient housing in Danang city, Vietnam". *Urban Climate, 12*, 85-103. doi:https://doi.org/10.1016/j.uclim.2015.01.002

[7] Tien Bui, D.; Duc, H.N.; Martínez-Álvarez, F.; Thao, N.T.T.; Hoa, P.V.; Dat, P.T.; Samui, P.; Costache, R. (2020). "A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area". *Science of The Total Environment, 701*,134413. doi:https://doi.org/10.1016/j.scitotenv.2019.134413

[8] Thi An, T.; Izuru, S.; Narumasa, T.; Raghavan, V.; Hanh, L.N.; Van An, N.; Long, N.V.; Thi Thuy, N.; Minh, T.P. (2022). "Flood vulnerability assessment at the local scale using remote sensing and GIS techniques: a case study in Da Nang City, Vietnam". *Journal of Water and Climate Change, 13 (9)*, 3217-3238. doi: https://doi.org/10.2166/wcc.2022.029

[9] Yuan, D. ; Wang H.; Wang C.; Yan C.; Xu L.; Zhang C.; Wang J.; Kou Y. (2024). "Characteristics of Urban Flood Resilience Evolution and Analysis of Influencing Factors: A Case Study of Yingtan City, China". *Water, 16 (6)*,834. doi: https://doi.org/10.3390/w16060834.