ARTICLE IN PRESS

Journal of Urban Management xxx (xxxx) xxx

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Journal of Urban Management

journal homepage: www.elsevier.com/locate/jum

Research Article

Machine learning-based suitability mapping for urban expansion: integrating environmental and infrastructural factors

Nguyen Thi Thuan [a,b], Nguyen Huu Ngu [a], Le Huu Ngoc Thanh [a], Duong Quoc Non [a], Srilert Chotpantarat [c,d], Nguyen Ngoc Thanh [a,*]

[a] *University of Agriculture and Forestry, Hue University, 102 Phung Hung Str, Hue City, 53000, Viet Nam*
[b] *Ho Chi Minh City University of Natural Resources and Environment, 236B Le Van Sy Str, Ho Chi Minh, 70000, Viet Nam*
[c] *Department of Geology, Faculty of Science, Chulalongkorn University, Bangkok, 10330, Thailand*
[d] *Center of Excellence in Environmental Innovation and Management (EnvIMM), Sustainable Research Institute, Chulalongkorn University, Thailand*

A R T I C L E I N F O

Keywords:
Built-up land
Mapping
Random forest
XGBoost
Generalized linear model
Vietnam

A B S T R A C T

Rapid urbanization in Ho Chi Minh City has increased conflicts between theory and practice in planned urban expansion. As a result, this study assessed the suitability of built-up land for supporting urban expansion decisions. This study contributes to machine learning-supported spatial planning by integrating environmental and infrastructural factors into suitability mapping. Using Generalized Linear Model (GLM), Extreme Gradient Boosting (XGBoost), and Random Forest (RF), we generated present-day built-up land suitability maps and compared them with the 2040 land use plan for urban expansion to assess their spatial consistency. A database of 6000 ground-truth points and eight influencing factors—including altitude, slope, soil type, proximity to waterbodies, proximity to residential areas, proximity to protected areas, proximity to roads, and topographic wetness index—was used to train and validate the models using a spatial 5-fold cross-validation. Model performance was assessed using the receiver operating characteristic curve (AUC), with RF (AUC = 0.958 ± 0.017) and XGBoost (AUC = 0.954 ± 0.018) outperforming GLM (AUC = 0.898 ± 0.023). Proximity to residential areas, waterbodies, and roads was an important factor in determining suitability for urban expansion. Most of the 2040 urban expansion was characterized by moderate to very low suitability levels. The study highlighted a substantial inconsistency between current environmental and infrastructural readiness and policy-designated future expansion areas in Ho Chi Minh City. Local authorities should therefore strengthen environmental and infrastructural preparedness before implementing urban expansion plans.

1. Introduction

Rapid urbanization is an alarming concern in many developing countries. This process involves converting agricultural, forest, and bare lands into built-up areas (Chen et al., 2016). Built-up land refers to areas covered by urban and rural settlements, industrial parks, infrastructure, and other developed facilities. This type of land use represents non-agricultural areas (Wu & Zhang, 2012). From a planning perspective, built-up land suitability is an important aspect of land-use planning that supports spatial decision-making (Guzman et al., 2020). However, the concept of "suitability" for built-up land has varied across previous studies. Some authors

* Corresponding author.
E-mail address: nguyenngocthanh@hueuni.edu.vn (N.N. Thanh).

https://doi.org/10.1016/j.jum.2026.100530
Received 7 October 2025; Received in revised form 4 April 2026; Accepted 12 June 2026
Available online xxxx

(Sihasale et al., 2023; Mallick et al., 2024; Qi et al., 2024) considered it a means of approving the expansion of built-up land, while others (Rimal et al., 2018; Bose & Chowdhury, 2020; Mohamed & Worku, 2020; Saxena & Jat, 2020) regarded it as a measure of urban growth. In this study, suitability for built-up land is defined as the potential to expand urban areas based on environmental and infrastructural conditions.

Although long-term land-use plans are commonly established to guide urban development, uncontrolled expansion of built-up land has frequently resulted in adverse environmental and land-use consequences, including agricultural land loss (Sadian & Shafizadeh-Moghadam, 2025), air pollution (Saxena, 2025), biodiversity decline (Liu et al., 2025), and the intensification of urban heat islands (Amir Siddique et al., 2021). These impacts show that urban expansion is both a development challenge and a spatial planning issue, highlighting the need for systematic assessment tools to identify locations that are environmentally and infrastructurally suitable for built-up development. In practice, mismatches between development pressures and land capability have frequently led to conflicts, as urban expansion directly competes with agricultural land conservation and environmental protection across contexts (Peerzado et al., 2019; de Jong et al., 2021; Wu et al., 2025). Ho Chi Minh City is a major economic center in Vietnam, with a high concentration of industrial, residential, and infrastructure zones. From 2000 to 2020, the built-up area expanded by approximately 20,000 ha (Tuan et al., 2025). The study area represents a low-lying, rapidly urbanizing coastal megacity, where environmental constraints and infrastructure readiness critically shape the feasibility of planned urban expansion. However, a spatial mismatch between actual development patterns and officially approved land-use plans has emerged in parts of the city, with expansion increasingly occurring in low-lying areas characterized by limited infrastructure. As a result, land suitability assessment is essential for more effective planning of built-up areas.

In the last decade, urban expansion has been conceptualized through urban sprawl and smart growth; these concepts provide a conceptual foundation for defining built-up land suitability. Urban sprawl is a type of development that spreads into suburban areas. It is typically low in density and widely dispersed. This type of growth leads to inefficient land use, increases infrastructure costs, and harms the environment (Brueckner, 2000; Chettry, 2023; Nechyba & Walsh, 2002). From a spatial planning perspective, inadequate consideration of environmental constraints and infrastructure capacity can lead to urban sprawl. In contrast, smart growth supports dense development. It values infrastructure efficiency, protects the environment, and ensures future sustainability. Smart growth discourages uncontrolled expansion. Instead, it promotes development in areas with good environmental conditions and sufficient infrastructure. Smart growth supports development in areas that meet land suitability criteria (Alexander & Tomalty, 2002; Litman, 2015; Mohammed et al., 2016). Therefore, built-up land suitability mapping can be regarded as an operationalization of smart growth principles, translating normative planning concepts into spatially explicit, data-driven decision-support tools.

Based on the above theoretical perspectives, the suitability assessment of built-up land can be considered in terms of environmental and infrastructural factors. Environmental factors such as altitude, slope, soil type, and the topographic wetness index (TWI) represent fundamental physical conditions that affect construction feasibility, flood risk, and environmental stability. These factors are especially important in low-lying coastal cities, where inappropriate land-use planning can increase exposure to climate-related hazards. Infrastructural factors—including proximity to residential areas, roads, waterbodies, and protected areas—are key considerations in spatial land-use planning. For example, proximity to residential areas and roads indicates infrastructure accessibility and service provision, whereas proximity to protected areas and waterbodies reflects environmental risk constraints. Moreover, their strong temporal variability and coarse spatial representation can increase uncertainty in modeling suitability. In this framework, socioeconomic factors are not included in the model because they primarily reflect development needs rather than the intrinsic conditions that determine land suitability. For these reasons, this study focuses on environmental and infrastructural factors that provide stable, spatially explicit information for suitability-based planning.

In land suitability assessment, geospatial technology has often been employed to prepare input factors (Jin et al., 2015; Ullah & Mansourian, 2016; Rudra et al., 2025). Nevertheless, the dataset of input factors is not consistent across studies. For instance, Mallick et al. (2024) identified 10 influencing factors, including elevation, distance from rivers and city centers, distance from service centers, land value, distance from railway and major roads, groundwater potential, land use/land cover, soil, and lithological structure. Thapa et al. (2025) focused on land cover, slope, terrain elevation, population density, proximity to infrastructure, proximity to brick factories, and flood sensitivity zones. The number of input factors depends on data availability and the characteristics of the study region. However, most influencing factors were derived from geographic information system and remote sensing. Land suitability maps can be generated once the geospatial data are defined.

Recent reports have applied various decision-making models to map urban expansion suitability. Among the most common are Multi-Criteria Decision Analysis (Ray et al., 2025), Multi-Influence Factor (Singh et al., 2021), Analytical Hierarchy Process (AHP) (Ismaeel & Kumar, 2023), Multi-Criteria Evaluation, and TOPSIS (Saleem et al., 2022); however, these models may not be suitable for big data applications (Auret & Aldrich, 2012). Although machine learning (ML) has been applied across diverse case studies, it has not yet been widely adopted in this field. Recently, Manna et al. (2025) compared the performance of Support Vector Machine (SVM), Artificial Neural Network (ANN), and Random Forest (RF) for assessing the suitability of built-up sites for urban redevelopment in India. Compared with decision-making models, ML has often produced stronger predictive results (Huang et al., 2020; Shikhteymour et al., 2023; Thanh et al., 2022). Huang et al. (2020) reported that ML models achieved areas under the receiver operating characteristic (ROC) curve of up to 0.86, outperforming general statistical and heuristic approaches (AUC <0.80) for landslide susceptibility mapping. Similarly, Thanh et al. (2022) showed that an RF model outperforms frequency-ratio and AHP-based approaches for groundwater potential mapping. In the context of flood risk assessment, Shikhteymour et al. (2023) reported superior predictive performance of ML models, with SVM achieving an AUC of 0.94 and the tree-based CART model reaching an AUC of 0.86. ML can analyze both linear and nonlinear relationships among input parameters. One challenge, however, is its demand for large datasets, which limits its usefulness in studies with limited data (Achu et al., 2020; Ngu et al., 2025). Additionally, an appropriate ML model for mapping built-up land suitability must be based on input data structure and format, interpretability, and feasibility. In built-up land suitability mapping, influencing factors exhibit high variability, including continuous variables (e.g., altitude, slope, distance to rivers, and residential areas) and categorical variables (e.g., land use/land cover and soil types).

RF, Generalized Linear Model (GLM), and Extreme Gradient Boosting (XGBoost), are relatively simple models that offer effective data processing capabilities and are straightforward to interpret. Meanwhile, other models—such as SVM, ANN, and K-Nearest Neighbors—have achieved high accuracy in certain cases but require more complex data processing. This disadvantage makes them less suitable for in-depth analysis in mapping built-up land suitability (Mountrakis et al., 2011; Martins et al., 2016; Linardatos et al., 2020; Sheykhmousa et al., 2020; Pinheiro et al., 2025). Based on these considerations, this study applied ML models to map the suitability of built-up land.

As a result, this study applied ML models to assess the current inherent suitability of built-up land based on present-day environmental and infrastructural conditions. An illustrative example was presented in Ho Chi Minh City, where output maps were evaluated using the ROC curve. The principal objectives of this study were to: (1) compare the performance of RF, XGBoost, and GLM in mapping suitability for built-up land; (2) analyze how environmental and infrastructural factors contribute to built-up land suitability under current conditions; and (3) assess the spatial consistency between the 2040 urban expansion plan and the present-day suitability for built-up land.

2. Methodologies

2.1. Study area

This study was conducted within the eastern zone of Ho Chi Minh City, Vietnam (106°41′49″ to 106°52′56″ E and 10°44′27″ to 10°53′55″ N, Fig. 1). It covers approximately 21,066.99 ha. About 60% of the study area is at elevations below 2 m above sea level (Phu et al., 2023). By 2023, the study area had a population of nearly 1.28 million. The climate is divided into the wet season (May–October) and the dry season (November–April). The annual precipitation in the area ranged from 1100 to 2000 mm/year (Phuong et al., 2019). The land surface temperature ranged from 30 to 34°C (Thi et al., 2025). The main land-use categories comprise built-up land (13, 511.61 ha), crop land (4023.83 ha), bare land (1.56 ha), and waterbodies (3529.99 ha), with an urbanization rate of approximately 24% between 2015 and 2020 (Quang et al., 2023). Built-up land accounts for approximately 87% of the total area in 2040. This urbanization process can result in conflicts between land use and environmental sustainability. Therefore, this region can serve as a test bed for evaluating the effectiveness of ML models in suitability assessment for urban expansion.

2.2. Data acquisition

Eight influencing factors—including altitude, slope, soil type, proximity to waterbodies, proximity to residential areas, proximity to protected areas, proximity to roads, and TWI—were first prepared to map suitability for built-up land. These factors were selected

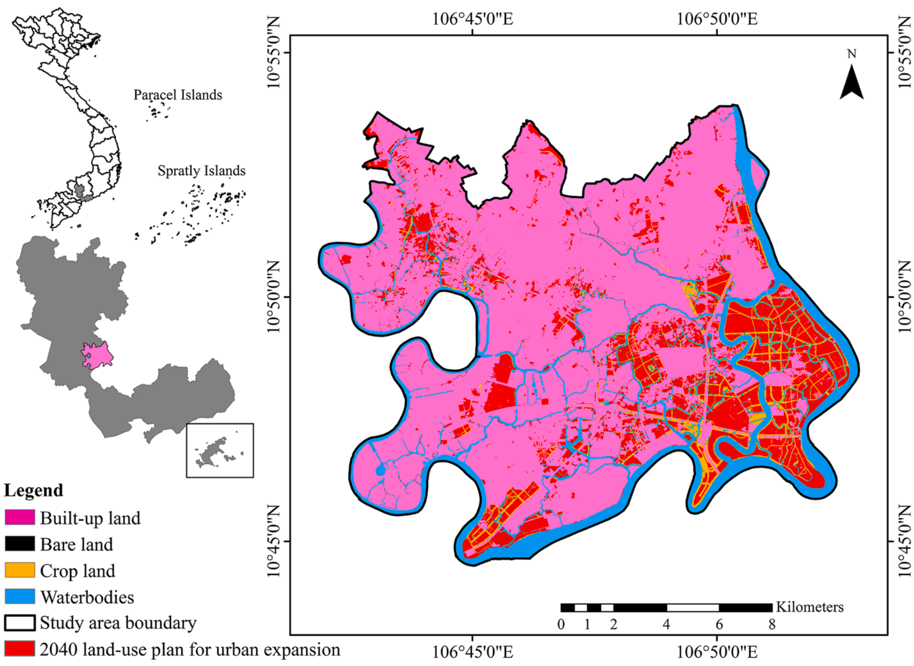

Fig. 1. Study area.

based on previous literatures (Ullah & Mansourian, 2016; Mallick et al., 2024; Parry et al., 2018; Rimal et al., 2018; Weldu & Deribew, 2016) and their availability for the study area. All factors were derived from existing and satellite data sources. ALOS PALSAR data were used for altitude and slope measurement (Fig. 2a and b). The soil type map was derived from the Ho Chi Minh City Agriculture and Environment Department (Fig. 2c). Proximity to waterbodies, residential areas, protected areas, and roads was derived from OpenStreetMap data. Waterbodies include natural features (rivers and streams) and artificial features (canals and ponds). Proximity to residential areas represents the impact of existing settlements and infrastructure on urban expansion. Proximity to protected areas was considered a regulatory constraint, as a greater distance from them generally implies fewer development restrictions and lower ecological conflict. In this study, hard planning constraints (e.g., protected areas, flood-prone zones, and steep terrain) are implicitly represented through proximity to protected areas. This approach allows the models to capture transitions in development potential rather than enforcing strict exclusion masks. Proximity to roads was used as a proxy for transportation accessibility. The "Euclidean distance" function in ArcMap 10.8 software was applied to create these distance maps (Fig. 2d–g). The TWI map was computed from slope and altitude layers (Pourali et al., 2016) (Fig. 2h). All influencing factors were standardized to a spatial resolution of 12.5 m using

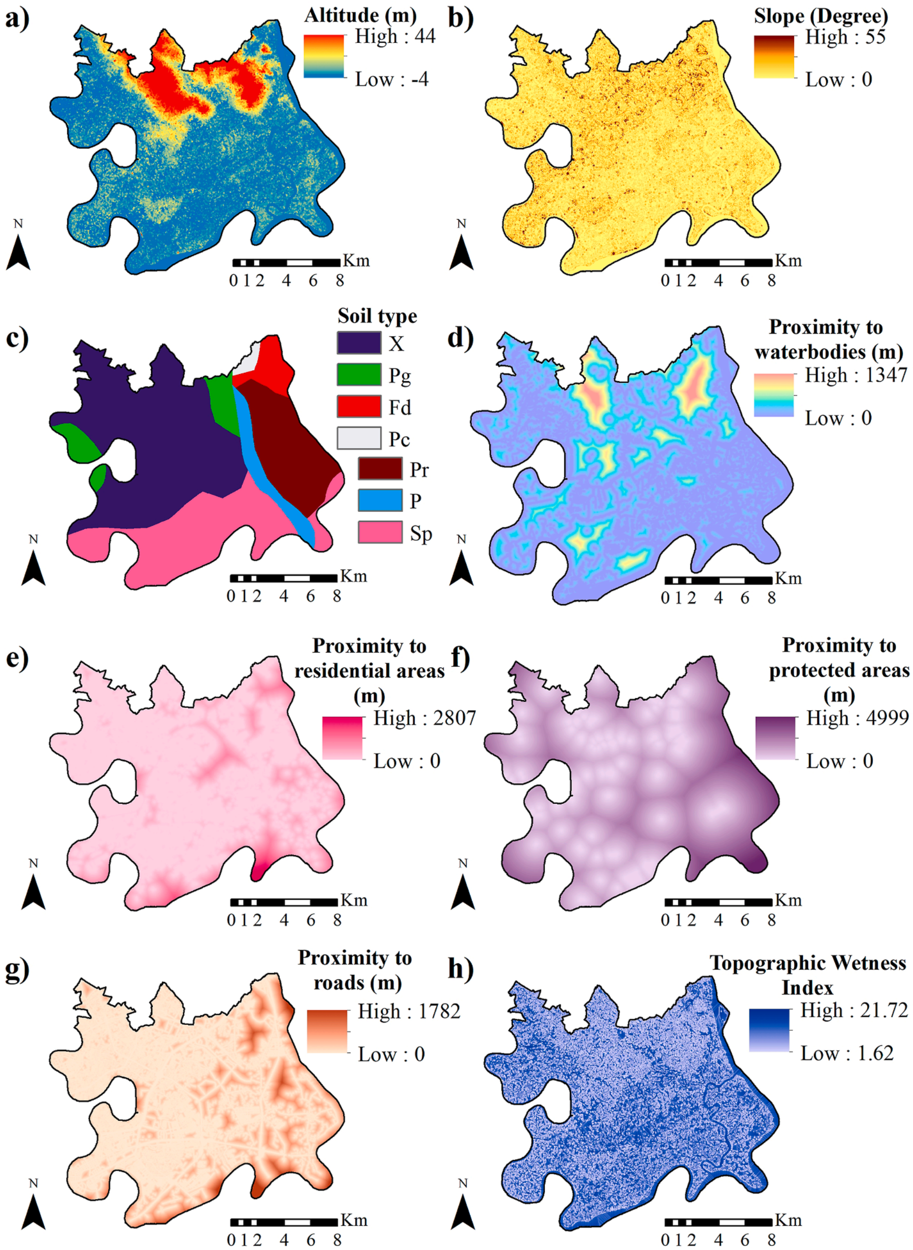

Fig. 2. Influencing factors: a) Altitude, b) Slope, c) Soil type, d) Proximity to waterbodies, e) Proximity to residential areas, f) Proximity to protected areas, g) Proximity to roads, h) Topographic wetness index.

the WGS 84 UTM Zone 48 N coordinate system. A concise overview of the influencing factors is provided in Table 1. Next, a stratified random sampling strategy was applied to design ground-truth points. A total of 6000 sample points were randomly generated from the land-use map. Given a spatial resolution of 12.5 m, the study area (21,066.99 ha) comprises approximately 1.35 million raster cells. The selected 6000 points represent approximately 0.44% of the total spatial population. These ground truths were obtained from the official land-use map. These labels were subsequently validated through a field survey. This process ensured reliable identification of built-up and non-built-up areas. Then, these sample points were coded as 0 for non-built-up areas and 1 for built-up areas. A field survey was subsequently conducted to verify the accuracy of these points. Next, we designed a dataset of 6000 points by associating each ground-truth point with the corresponding values of the eight influencing factors. In the models, the influencing factors were the input variables, and the ground-truth points were the response variable. Soil type was defined as a categorical variable, while the other variables were defined as continuous. All preprocessing steps were completed before data partitioning from raster standardization to feature extraction at sample locations. A spatial 5-fold cross-validation approach was employed, in which sample points were first grouped into spatial blocks of 2000 m based on their coordinates. These blocks were randomly assigned to five folds, ensuring that all points within a given block were allocated to the same fold. In each iteration, four folds (80% of the data) were used for model training, while the remaining fold (20%) served as the validation set. This process was repeated five times, such that each fold was used once for the validation set.

2.3. *Calculating suitability index for built-up land*

2.3.1. *RF*

RF, a bagging-based ML algorithm, employs decision trees to address regression and classification problems (Chen et al., 2017). According to previous reports (Taalab et al., 2018; Wu et al., 2021; Thanh et al., 2022), RF performs well on large datasets, particularly those exhibiting nonlinear relationships and complex interactions. For each decision tree, a bootstrap subset of the training data is used to train the model. By selecting a random subset of predictors at each split, RF introduces dual randomization, which lowers tree correlations and, as a result, enhances generalization while reducing overfitting. For the present analysis, RF was run in RStudio with the "randomForest" package. For our RF configuration, 500 decision trees were generated (ntree = 500). At each split, a random subset of predictor variables (mtry = 2) was considered. Each tree was trained on a bagging sample of the training dataset. No limit was set on the number of nodes. Soil type was a categorical variable. It was handled directly as a "factor" within the R environment. The RF model was implemented using default hyperparameter settings to ensure transparency and comparability with other models. Variable importance was assessed using a permutation-based approach. During training, the "Mean Decrease in Accuracy" index was estimated for each variable based on out-of-bag permutation samples (i.e., observations not included in the bootstrap sample). This index was then normalized to express the percentage contribution of each influencing factor to the overall ranking. After training, the RF model was applied across influencing factor layers to generate a continuous probability raster, with each pixel containing a value between 0 and 1.

2.3.2. *XGBoost*

XGBoost, a boosting-based ML algorithm, is based on the gradient principle (Chen et al., 2015). In contrast to RF, which generates trees simultaneously, XGBoost constructs them sequentially. In XGBoost, each new decision tree is built to adjust for the residual errors of earlier trees. A key strength of XGBoost lies in its ability to control complexity and alleviate overfitting through its sequential boosting process. In this study, XGBoost was implemented in RStudio using the "xgboost" package. Because XGBoost requires numeric input data, the soil type factor was one-hot encoded using the "model.matrix" function to create the feature matrix. This encoding ensures appropriate treatment of the nominal soil type factor. Then, we set up the XGBoost model to run 100 boosting iterations (nrounds = 100) during training. For the learning task, the objective function "binary:logistic" was chosen, enabling binary classification. The XGBoost model was trained on the training dataset without early stopping, ensuring that every boosting iteration was carried out. At the same time, other hyperparameters remained at their default values (eta = 0.3, max_depth = 6, subsample = 1) in the

Table 1
Overview of influencing factors.

No	Influencing factors	Data format	Code	Subgroups	Property
1	Altitude	Continuous	–	–	-4–44 m
2	Slope	Continuous	–	–	0°-55°
3	Soil type	Categorical	1	Acrisols (X)	9297.65
			2	Gleyic fluvisols (Pg)	1263.11 ha
			3	Rhodic ferralsols (Fd)	831.85 ha
			4	Dystric fluvisols (Pc)	198.55 ha
			5	Cambic fluvisols (Pr)	3215.91 ha
			6	Eutric fluvisols (P)	1028.29 ha
			7	Proto-thionic gleysols (Sp)	5231.63 ha
4	Proximity to waterbodies	Continuous	–	–	0–1347 m
5	Proximity to residential areas	Continuous	–	–	0–2807 m
6	Proximity to protected areas	Continuous	–	–	0–4999 m
7	Proximity to roads	Continuous	–	–	0–1782 m
8	TWI	Continuous	–	–	1.62–21.72

"xgboost" package. The importance of influencing factors was determined using a permutation-based approach. The importance value was normalized into a percentage to estimate the impact of each influencing factor on the suitability index prediction. After training, the XGBoost model was applied to all layers of influencing factors to generate a continuous probability raster, with each pixel containing a value between 0 and 1.

2.3.3. GLM

GLM, a logistic regression-based ML algorithm, employs composite link functions and probability distributions (Ghanbarian et al., 2019). In GLM, event probabilities are mapped using the logit function, which converts them to a suitable scale. Through this transformation, the logarithmic odds of the target factor are a weighted sum of predictors (Ahlawat, 2025). In addition, GLM offers flexibility, interpretability, and the ability to accommodate various data types. In this study, GLM was implemented in RStudio using the "glm" package with the argument "family = binomial (link = "logit")". The GLM was fitted based on the training dataset without preliminary variable screening. Soil type was treated as a categorical factor variable. This specification allows estimation of class-specific effects without assuming a numerical structure. The regression coefficients were estimated by maximum likelihood. The importance of influencing factors was determined using a permutation-based approach. After training, the GLM was applied to all layers of influencing factors to generate a continuous probability raster, with each pixel containing a value between 0 and 1.

2.3.4. Validation and sensitivity analysis

Model validation is an essential procedure for assessing and comparing the predictive capability of RF, XGBoost, and GLM. Model performance was quantified through the ROC curve and its corresponding AUC value. The predicted probability for class "built-up" was extracted from each model. For the RF, we used the decision tree vote shares. The logistic probability from the boosting process was used for the XGBoost, and the probability from the logistic function was applied to GLM. The AUC score ranges from 0 to 1, with higher values indicating better model performance.

For the sensitivity analysis, a partial dependence plot was generated to examine the role of each influencing factor in determining the suitability index using the "pdp" and "ggplot 2" libraries.

2.4. Assessment of present-day built-up land suitability for the 2040 urban expansion

To assess the consistency between current land readiness and planned future development, the present-day built-up land suitability was compared with the 2040 urban expansion land-use planning data from the Ho Chi Minh City Agriculture and Environment Department. The map was designated as a raster with a resolution of 12.5 m, reprojected to WGS 84 UTM Zone 48 N, and clipped to the study boundary. We then conducted a statistical analysis of present-day suitability levels and spatial extent within these planned regions. For analytical visualization and improved decision-making, suitability index maps were categorized into five classes using Equal Interval, Quantile, Natural Breaks (Jenks) classification schemes. The Akaike Information Criterion (AIC) metric was applied to assess sensitivity and minimize additional uncertainty in map interpretation arising from these classification approaches. Lower AIC values indicate better classification performance (Moreira et al., 2021). AIC was computed as follows:

$$AIC = -2\sum_{k=1}^{m}\sum_{i\in G_k} x_i \log(\widehat{q}_k) + 2(m-1) \tag{1}$$

$$\widehat{q}_k = \sum_{i\in G_k}\frac{x_i}{Xn_k} \tag{2}$$

where, x_i is the suitable value of pixel i, m is the number of classes, G_k is the kth class, X is the total sum of suitability values, and n_k is the number of pixels in class G_k.

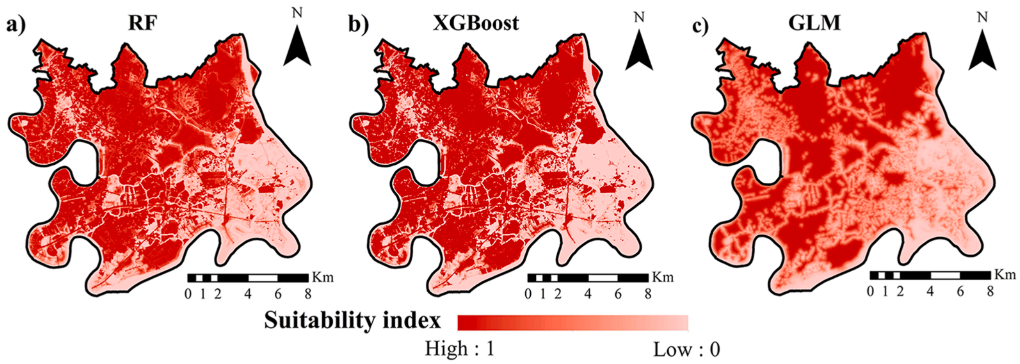

Fig. 3. Suitability maps for built-up land.

3. Results

3.1. Spatial distribution of suitability for built-up land

In our case study, MLs—including RF, XGBoost, and GLM—were applied to estimate the suitability index for built-up land in the study area. Fig. 3a–c shows that zones with high suitability for built-up land were concentrated in the urbanized zones, while zones with low suitability were concentrated in the southeastern portions. However, the models differed significantly in their spatial patterns. The RF and XGBoost presented similar results. The GLM depicted zones with high suitability in the urban core. Table 2 illustrates the contribution of the influencing factors. In the RF model, the contribution level to built-up land suitability of proximity to waterbodies was the highest (44.03%). It was followed by proximity to residential areas (38.08%) and proximity to roads (12.20%). Soil type, proximity to protected areas, altitude, slope, and TWI ranked next in contribution levels at 2.25%, 1.73%, 1.47%, 0.21%, and 0.03%, respectively. For the XGBoost model, the highest contribution (45.27%) was found near waterbodies, followed by residential areas (35.04%). Meanwhile, the figures for proximity to roads and proximity to protected areas were 12.19% and 2.97%, respectively. The other influencing factors had a low impact. Regarding the GLM model, proximity to waterbodies accounted for 77.03% of the contribution.

3.2. Model performance evaluation

The predictive performance of the models was validated by the spatial 5-fold cross-validation. The AUC values of RF and XGBoost were 0.958 ± 0.017 and 0.954 ± 0.018, respectively, whereas the AUC of GLM was 0.898 ± 0.023 (Fig. 4). These results indicate that RF and XGBoost are better than GLM.

3.3. Sensitivity analysis of influencing factors

Fig. 5 provides the association between the suitability index and the explanatory factors. Altitude had an inverse relationship with the suitability index (Fig. 5a). High suitability zones for built-up land were generally found at low altitudes. The suitability index remained relatively constant with slopes between 0° and 2° (Fig. 5b). There were notable differences in suitability among soil types, with the highest values observed in cambic fluvisols, eutric fluvisols, and proto-thionic gleysols (Fig. 5c). Proximity to waterbodies strongly influenced the suitability index. The suitability index decreased from approximately 0.75 to 0.20 when proximity to waterbodies increased from 0 to 350 m (Fig. 5d). Proximity to residential areas was more suitable in locations immediately adjacent to settlements (Fig. 5e). The suitability index showed an increasing trend as proximity to protected areas increased (Fig. 5f). For proximity to roads, the suitability index remained stable at around 0.55 (Fig. 5g). The relationship between TWI and the suitability index remained around 0.50 across different levels (Fig. 5h).

3.4. Present-day built-up land suitability within areas designated for urban expansion in 2040

Fig. 6 depicts the present-day suitability levels for built-up land in the planned regions using Equal Interval, Quantile, Natural Breaks (Jenks) classification schemes. The majority of the planned areas were classified as low and very low in terms of current environmental and infrastructural readiness (Fig. 7). In the RF model, using the Equal Interval classification, 82.80% of the planned areas were categorized as very low or low suitability, followed by moderate suitability (7.88%) and high or very high suitability (9.32%). Under the Quantile classification, these proportions were 40.54%, 20.03%, and 39.43%, respectively, while under Natural Breaks they were 72.16%, 12.97%, and 14.87%, respectively. A similar pattern was observed for the XGBoost model, very low and low suitability dominated, accounting for 84.99% (Equal Interval), 44.91% (Quantile) and 79.59% (Natural Breaks). Moderate suitability was limited (4.87%, 19.64%, and 8.15%, respectively). High and very high suitability areas remained low under Equal Interval (10.14%), and Natural Breaks (12.26%), but increased substantially under Quantile (35.45%). In the GLM model, most planned areas were also categorized as very low or low suitability, comprising 69.74% (Equal Interval), 41.34% (Quantile), and 60.56% (Natural Breaks). Moderate suitability accounted for 17.29%, 20.41%, and 21.70%, respectively. High and very high suitability classes represented 12.97% under Equal Interval, 38.25% under Quantile, and only 17.74% under Natural Breaks. Table 3 and Fig. 8 present the

Table 2
Contribution of influencing factors to the present-day built-up land suitability index.

Models	RF	XGBoost	GLM
Altitude	1.47	1.40	1.04
Slope	0.21	0.37	0.13
Soil type	2.25	2.39	5.01
Proximity to waterbodies	44.03	45.27	77.03
Proximity to residential areas	38.08	35.04	7.23
Proximity to protected areas	1.73	2.97	2.66
Proximity to roads	12.20	12.19	6.78
TWI	0.03	0.37	0.12
Total	**100.00**	**100.00**	**100.00**

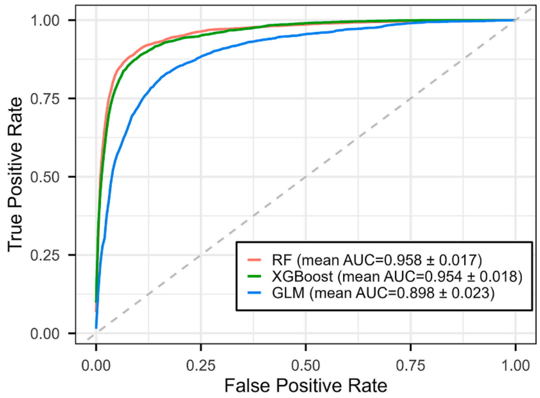

Fig. 4. Model validation.

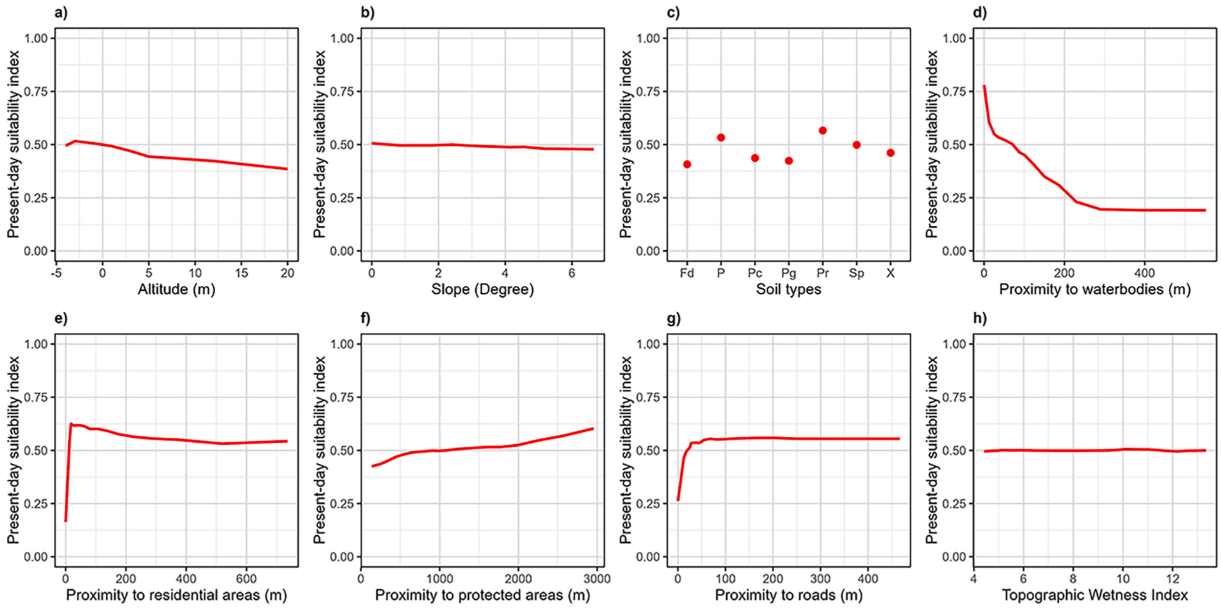

Fig. 5. Sensitivity of influencing factors on suitability for urban expansion.

distributions of present-day suitability probability within the 2040 planned urban expansion areas. Sensitivity analysis using the AIC indicated that the Natural Breaks performed best among the three schemes across the RF, XGBoost, and GLM models (Table 4).

4. Discussion

As highlighted in the introduction, delineating suitable zones for built-up land is a key strategy for land-use planning. Several reports (Saleem et al., 2022; Liladhar Rane et al., 2024; Manna et al., 2025) have revealed that alternative approaches produce distinct outcomes. This work was designed to compute a suitability index for built-up land based on eight influencing factors and ML algorithms. This study found that the output maps of RF and XGBoost differed from those of GLM (Fig. 3), which can be explained by the decision-tree-based approach of RF and XGBoost, whereas GLM is based on the linearity assumption. Compared with the 2040 land-use planning for built-up land (Fig. 1), most of the planned expansion areas fell into very low, low, and moderate present-day suitability classes in the models (Fig. 6). This inconsistency may reflect differences in planning logic and temporal perspectives. The 2040 land-use planning for built-up land is a policy instrument, in which designated expansion areas often function as long-term development reserves aligned with socio-economic and infrastructure investment targets.

In contrast, the ML-based suitability assessment evaluated the potential for urban expansion under current environmental and infrastructural conditions. As a result, planned regions that lacked developed infrastructure—particularly agricultural areas with limited road networks and residential facilities—were classified as low suitability by the models. Previous studies have reported land mismatches associated with urban growth and spatial planning (Saha & Roy, 2021; Scott et al., 2013), showing that such discrepancies

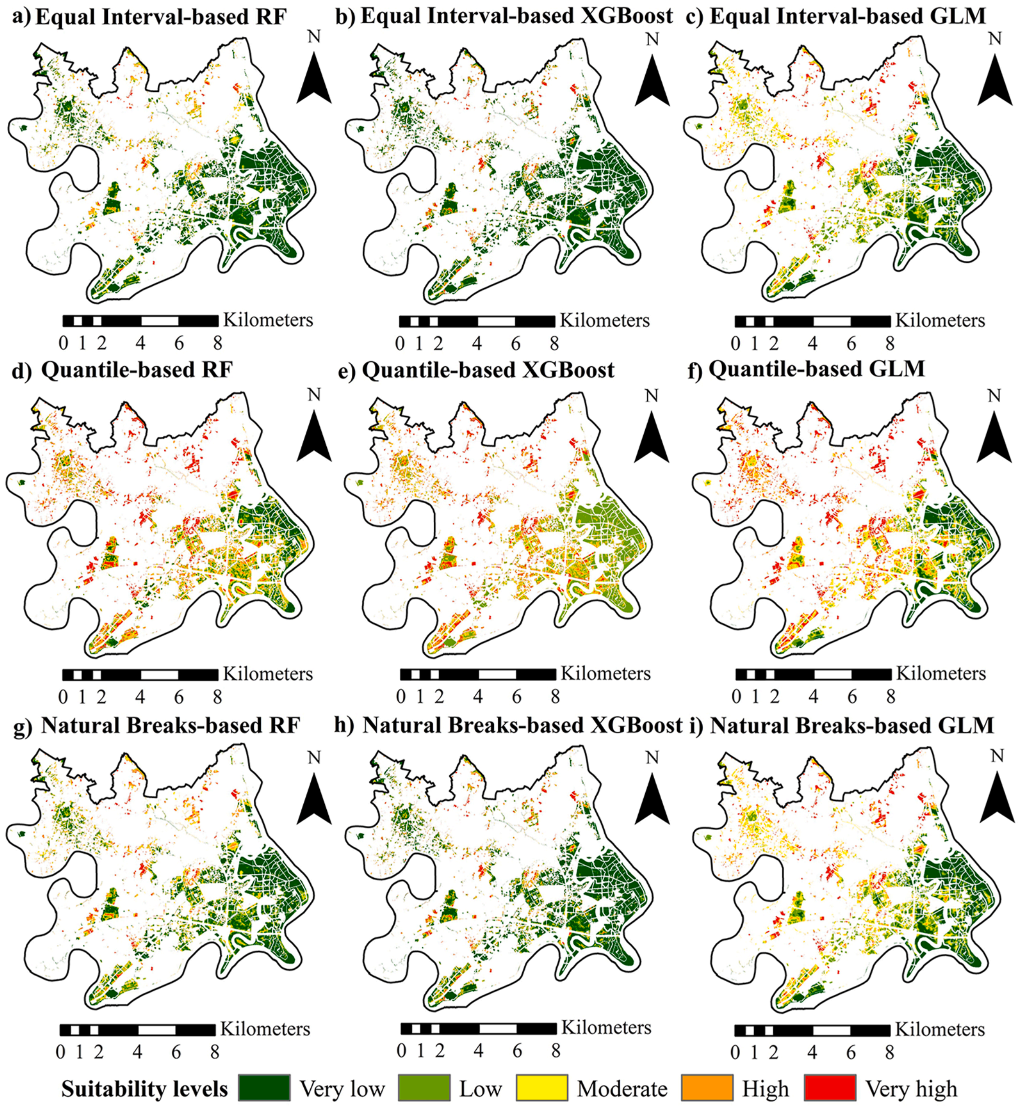

Fig. 6. Present-day suitability maps for urban expansion in 2040.

are widespread in suburban and peri-urban areas. This mismatch can be interpreted as a gap between planning intentions and current conditions, necessitating substantial infrastructure upgrades to realize the planned development.

It is important to emphasize that the models developed in this study are not intended to propose optimal locations for future urban development. Instead, they estimate present-day built-up land suitability under the assumption that current environmental and infrastructural conditions remain continuous. Consequently, these models are more suitable for diagnosing existing spatial patterns and evaluating the consistency between current conditions and planned land-use targets, rather than for designing future development scenarios that would require explicit consideration of policy objectives, legal constraints, and planned infrastructure investments.

From a management perspective, this misalignment serves as an early warning of potential implementation risks when urban expansion is driven primarily by policy objectives without sufficient consideration of environmental and infrastructural constraints. In such cases, developing areas with low suitability and no prior infrastructure upgrades are likely to incur higher long-term costs for drainage, flood mitigation, transportation, and environmental remediation. Moreover, such expansion may exacerbate urban sprawl and reduce planning efficiency by dispersing development into areas that are not yet functionally prepared. To reduce potential environmental and fiscal risks, urban managers should incorporate suitability-based evidence into planning implementation, particularly when prioritizing infrastructure investment and sequencing development. These findings suggest that suitability-based assessments should be incorporated into urban governance frameworks to support risk-aware and cost-efficient implementation of long-term development plans.

The RF and XGBoost achieved higher AUC values than the GLM (Fig. 4). This evidence indicates that ensemble MLs were more efficient at capturing the complex interactions between influencing factors and the suitability index than the linear ML. The RF

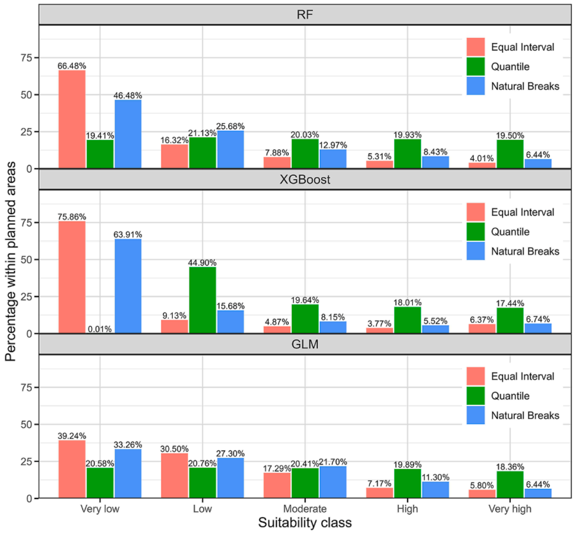

Fig. 7. Present-day suitability for urban expansion in 2040.

Table 3
Statistics of predicted suitability probabilities within the 2040 planned urban expansion areas.

Model	n pixels	Mean	Median	Q1	Q3	IQR	Min	Max	SD
RF	244,023	0.205	0.112	0.038	0.278	0.240	0	1	0.233
XGBoost	244,023	0.163	0.030	0.002	0.188	0.186	0	1	0.265
GLM	244,023	0.327	0.274	0.129	0.456	0.327	0	1	0.249

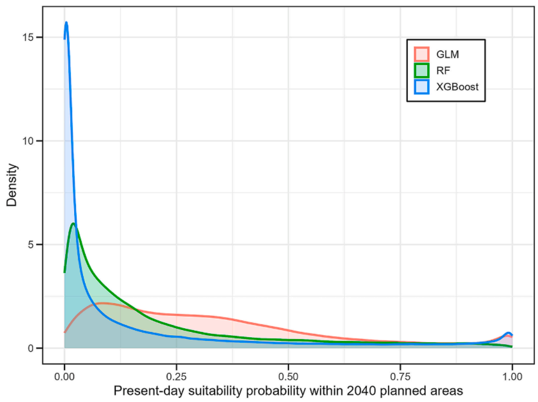

Fig. 8. Density distributions of suitability index within the 2040 planned urban areas.

Table 4
AIC of classification schemes.

Classification schemes	RF	XGBoost	GLM
Equal Interval	1191768.00	915279.08	1948054.00
Quantile	1187658.00	909211.20	1948664.00
Natural Breaks	1187529.00	909014.60	1947817.00

leveraged a bootstrap mechanism to reduce variance, while the XGBoost used boosting to minimize errors. In contrast, the GLM models the logit of the suitability index as linearly dependent on the influencing factors, potentially failing to account for complex nonlinear interactions. The result was consistent with previous studies. Kamusoko and Gamba (2015) reported that the RF was more reliable than the GLM in simulating urban growth. Another finding was that models using high-resolution input data can improve predictive performance. In contrast to Sun et al. (2019) and Virtriana et al. (2025), who used input data at 30 m resolution, this study achieved higher performance with 12.5 m-resolution input data, suggesting that the models were sensitive to input data resolution. Among the classification schemes, Natural Breaks demonstrated superior performance compared with Quantile and Equal Interval, consistent with previous findings by Akay and Baduna Koçyiğit (2024). This may be attributed to the way each method partitions the data. Equal Interval divides the data range into classes with equal width, whilst Quantile assigns an equal number of observations into each class. In contrast, Natural Breaks identifies class boundaries according to the inherent grouping of values, maximizing between-class differences while minimizing within-class variance.

In the RF, XGBoost, and GLM models, key influencing factors, particularly proximity to residential areas, waterbodies, and roads, explained 94.31%, 92.50%, and 91.04% of the variation of the suitability index, respectively (Table 2). Our findings align with previous studies by Liu et al. (2014) and Utami et al. (2018), which showed that proximity to infrastructure and hydrological systems has a significant influence on urban development. However, the contributions of individual factors to the present-day built-up land suitability index varied across models. These discrepancies reflect fundamental differences in how linear (GLM) and tree-based approaches (RF and XGBoost) handle mixed-type predictors and complex interactions, with the latter providing a more balanced representation of factor importance.

Regarding the relationship between the suitability index and influencing factors (Fig. 5), altitude showed an inverse correlation. Our observation indicated that the suitability index decreased as altitude increased; consistent with previous findings of Luan et al. (2021), which showed that lower-altitude zones are more favorable for urban development. For slope, the suitability index for built-up land remained stable at 0-2° because the study area's terrain was relatively flat. The findings of Saxena (2025) also indicated that zones with low slopes are generally favorable for urban development, and vice versa. Regarding soil type, the suitability index of Cambic fluvisols, Eutric fluvisols, and Proto-thionic gleysols was higher than that of other soil types. In contrast, Deliry and Uyguçgil (2020) reported that alluvial materials were unsuitable for urban development due to their susceptibility to volumetric changes under moisture variations. A possible explanation for our case study is that Cambic fluvisols, Eutric fluvisols and Proto-thionic gleysols were mainly distributed in flat terrain. Moreover, these zones already contained built-up land. The explanations imply that site-specific conditions are more important than soil type's intrinsic properties in determining suitability for urban development. For proximity to waterbodies, the suitability index decreased with increasing distance, indicating that proximity to rivers, lakes, or ponds was a priority for citizens. This result supports previous studies (Jandaghian & Colombo, 2024; Zhou et al., 2022) that have highlighted the important role of waterbodies in regulating the urban microclimate and mitigating urban heat islands. From a spatial planning perspective, this finding implies that urban expansion near waterbodies requires careful integration of flood mitigation and drainage infrastructure. Although proximity to waterbodies enhances urban livability, unregulated development in these areas may increase long-term costs related to flood control and environmental management, particularly in low-lying coastal cities such as Ho Chi Minh City. This study indicated that proximity to residential areas was associated with a higher suitability index for built-up land. This finding was consistent with Ullah and Mansourian (2016), who reported that zones near residential areas were moderately to highly suitable for built-up land because they benefit from existing infrastructure and public services, suggesting that lands adjacent to or surrounding residential areas offer many advantages for urban expansion. This study supports infill and edge-expansion strategies, whereby urban growth is prioritized adjacent to existing residential areas to maximize the use of existing infrastructure and minimize the costs of service extension. Our study confirmed that the suitability index and proximity to protected areas were positively correlated. A greater distance from protected areas indicated a greater suitability. This correlation can be explained by the fact that maintaining an adequate distance helps protected areas sustain their ecological functions (De La Fuente et al., 2020). Surprisingly, proximity to roads was found to have only a threshold effect. The suitability index stabilized around 0.55. This finding is contrary to that of Sihasale et al. (2023), who reported that suitability for built-up land decreased with increasing proximity to roads. This inconsistency may be due to the relatively uniform transport network in the study area, which ensures accessibility even as distance to roads increases. The threshold effect observed for proximity to roads suggests that beyond a basic level of accessibility, additional proximity does not significantly enhance suitability; this indicates that in areas with relatively uniform transport networks, road expansion alone may not be an effective driver of sustainable urban development; instead, coordinated investment in drainage systems, public services, and land-use control may play a more decisive role. According to Xi et al. (2018), TWI also affects urban development via flood risk management. They stated that higher TWI values correspond to greater risk. However, our observation showed the suitability index for built-up land was approximately 0.50 across all TWI values in the study area. These differences can be partly explained by the flat terrain characteristics within the study region.

This study used the RF, XGBoost, and GLM algorithms, along with multi-source raster datasets (i.e., altitude, slope, soil type, proximity to waterbodies, proximity to residential areas, proximity to protected areas, proximity to roads, and TWI). Multi-source data were uniformly processed to minimize spatial bias and ensure comparability across models. An outstanding feature of this study was its ability to reveal nonlinear relationships and complex interactions among influencing factors in mapping built-up land suitability. However, this study also had some drawbacks. Firstly, the input variables represented environmental and infrastructural factors. At the same time, economic and social conditions (e.g., population density, land price, and development pressure) were not integrated, which may reduce the comprehensiveness of built-up land suitability mapping. As a result, the derived suitability index mainly reflects physical and infrastructural feasibility rather than socio-economic factors. Secondly, the database in this study was collected at a single point in time; therefore, it did not reflect the fluctuations over time or different development scenarios. Consequently, the results represent a static assessment and may not fully reflect future changes in infrastructure provision or land-use dynamics. Thirdly, the model outcomes may not be directly generalized to mountainous areas, as this study was conducted on flat terrain. As a result, future studies are encouraged to incorporate spatial, temporal, and socio-economic data to test the generalizability of built-up land suitability maps.

5. Conclusion

This study applied ML models—including RF, XGBoost, and GLM—to map present-day built-up land suitability for expansion in Ho Chi Minh City based on environmental and infrastructural factors. Our results demonstrated that RF and XGBoost outperformed GLM, and that proximity to residential areas, waterbodies, and roads were the most influential factors in estimating the suitability index. Most areas designated for urban expansion in the 2040 plan ranged from very low to moderate present-day suitability, mainly concentrated in areas with underdeveloped infrastructure. Our study provides a policy message that urban expansion in Ho Chi Minh City, particularly where characterized by low suitability, can entail high long-term costs for transportation infrastructure, drainage networks, flood mitigation, and environmental management. As a result, the findings of this study can provide evidence for making informed decisions in urban planning in Ho Chi Minh City. By integrating environmental and infrastructural factors, ML-supported spatial planning research enables urban authorities to allocate infrastructure investments more efficiently in each development phase in the future. Although the current study primarily reflects environmental and infrastructural suitability, future work should integrate socio-economic factors and temporal dynamics to better capture planning scenarios. We also recommend applying the framework across diverse topographical settings to evaluate its applicability in rapidly developing cities.

CRediT authorship contribution statement

Nguyen Thi Thuan: Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft. **Nguyen Huu Ngu:** Writing – review & editing. **Le Huu Ngoc Thanh:** Writing – review & editing. **Duong Quoc Non:** Writing – review & editing. **Srilert Chotpantarat:** Writing – review & editing. **Nguyen Ngoc Thanh:** Conceptualization, Data curation, Formal analysis, Investigation, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors acknowledge the partial support of Hue University under the Core Research Program, Grant No. NCTB.DHH.2026.05.

References

Achu, A. L., Aju, C. D., & Reghunath, R. (2020). Spatial modelling of shallow landslide susceptibility: A study from the southern Western ghats region of Kerala, India. *Annals of GIS, 26*, 113–131. https://doi.org/10.1080/19475683.2020.1758207

Ahlawat, S. (2025). Generalized linear model. In *Statistical quantitative methods in finance: From theory to quantitative portfolio management* (pp. 41–91). Springer.

Akay, H., & Baduna Koçyiğit, M. (2024). Investigation of flood hazard susceptibility using various distance measures in technique for order preference by similarity to ideal solution. *Applied Sciences, 14*, 7023. https://doi.org/10.3390/app14167023

Alexander, D., & Tomalty, R. (2002). Smart growth and sustainable development: Challenges, solutions and policy directions. *Local Environment, 7*, 397–409. https://doi.org/10.1080/1354983022000027578

Amir Siddique, M., Wang, Y., Xu, N., Ullah, N., & Zeng, P. (2021). The spatiotemporal implications of urbanization for urban heat islands in Beijing: A predictive approach based on CA–Markov modeling (2004–2050). *Remote Sensing, 13*, 4697. https://doi.org/10.3390/rs13224697

Auret, L., & Aldrich, C. (2012). Interpretation of nonlinear relationships between process variables by use of random forests. *Minerals Engineering, 35*, 27–42. https://doi.org/10.1016/j.mineng.2012.05.008

Bose, A., & Chowdhury, I. R. (2020). Monitoring and modeling of spatio-temporal urban expansion and land-use/land-cover change using markov chain model: A case study in Siliguri Metropolitan area, West Bengal, India. *Modeling Earth Systems and Environment, 6*, 2235–2249. https://doi.org/10.1007/s40808-020-00842-6

Brueckner, J. K. (2000). Urban sprawl: Diagnosis and remedies. *International Regional Science Review, 23*, 160–171.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., & Zhou, T. (2015). Xgboost: Extreme gradient boosting. *R package version 0.4-2 1*, 1–4.

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., Duan, Z., & Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena, 151*, 147–160. https://doi.org/10.1016/j.catena.2016.11.032

Chen, Y., Chen, Z., Xu, G., & Tian, Z. (2016). Built-up land efficiency in urban China: Insights from the general land use plan (2006–2020). *Habitat International, 51*, 31–38. https://doi.org/10.1016/j.habitatint.2015.10.014

Chettry, V. (2023). A critical review of urban sprawl studies. *Journal of Geovisualization and Spatial Analysis, 7*, 28. https://doi.org/10.1007/s41651-023-00158-w

de Jong, L., De Bruin, S., Knoop, J., & van Vliet, J. (2021). Understanding land-use change conflict: A systematic review of case studies. *Journal of Land Use Science, 16*, 223–239. https://doi.org/10.1080/1747423X.2021.1933226

De La Fuente, B., Bertzky, B., Delli, G., Mandrici, A., Conti, M., Florczyk, A. J., Freire, S., Schiavina, M., Bastin, L., & Dubois, G. (2020). Built-up areas within and around protected areas: Global patterns and 40-year trends. *Global Ecology and Conservation, 24*, Article e01291. https://doi.org/10.1016/j.gecco.2020.e01291

Deliry, S.İ., & Uyguçgil, H. (2020). GIS-based land suitability analysis for sustainable urban development: A case study in Eskisehir, Turkey. *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi, 20*, 634–650.

Ghanbarian, G., Raoufat, M. R., Pourghasemi, H. R., & Safaeian, R. (2019). Habitat suitability mapping of Artemisia aucheri boiss based on the GLM model in R. *Spatial modeling in GIS and R for Earth and environmental sciences* (pp. 213–227). https://doi.org/10.1016/B978-0-12-815226-3.00009-0

Guzman, L. A., Escobar, F., Peña, J., & Cardona, R. (2020). A cellular automata-based land-use model as an integrated spatial decision support system for urban planning in developing cities: The case of the Bogotá region. *Land Use Policy, 92*, Article 104445. https://doi.org/10.1016/j.landusepol.2019.104445

Huang, F., Cao, Z., Guo, J., Jiang, S.-H., Li, S., & Guo, Z. (2020). Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *Catena, 191*, Article 104580. https://doi.org/10.1016/j.catena.2020.104580

Ismaeel, W. A., & Kumar, J. S. (2023). In *Analytical hierarchy process for land suitability analysis of urban growth in latakia, Syria* (pp. 191–201). https://doi.org/10.1007/978-981-99-6229-7_16

Jandaghian, Z., & Colombo, A. (2024). The role of water bodies in climate regulation: Insights from recent studies on urban heat island mitigation. *Buildings, 14*, 2945. https://doi.org/10.3390/buildings14092945

Jin, G., Li, Z., Lin, Q., Shi, C., Liu, B., & Yao, L. (2015). Land use suitability assessment in low-slope hilly regions under the impact of urbanization in Yunnan, China. *Advances in Meteorology, 2015*, Article 848795. https://doi.org/10.1155/2015/848795

Kamusoko, C., & Gamba, J. (2015). Simulating urban growth using a random forest-cellular automata (RF-CA) model. *ISPRS International Journal of Geo-Information, 4*, 447–470. https://doi.org/10.3390/ijgi4020447

Liladhar Rane, N., Purushottam Choudhary, S., Saha, A., Srivastava, A., Pande, C. B., Alshehri, F., Roy, R., Katipoğlu, O. M., & Abdo, H. G. (2024). Delineation of environmentally sustainable urban settlement using GIS-Based MIF and AHP techniques. *Geocarto International, 39*, Article 2335249. https://doi.org/10.1080/10106049.2024.2335249

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. *Entropy, 23*, 18. https://doi.org/10.3390/e23010018

Litman, T. (2015). *Understanding smart growth savings*. Canada: Victoria Transport Policy Institute Victoria, BC.

Liu, N., Liu, Z., & Wu, Y. (2025). Direct and indirect impacts of urbanization on biodiversity across the world's cities. *Remote Sensing, 17*, 956. https://doi.org/10.3390/rs17060956

Liu, R., Zhang, K., Zhang, Z., & Borthwick, A. G. L. (2014). Land-use suitability analysis for urban development in Beijing. *Journal of Environmental Management, 145*, 170–179. https://doi.org/10.1016/j.jenvman.2014.06.020

Luan, C., Liu, R., & Peng, S. (2021). Land-use suitability assessment for urban development using a GIS-based soft computing approach: A case study of Ili Valley, China. *Ecological Indicators, 123*, Article 107333. https://doi.org/10.1016/j.ecolind.2020.107333

Mallick, S. K., Rudra, S., & Maity, B. (2024). Land suitability assessment for urban built-up development of a city in the Eastern Himalayan foothills: A study towards urban sustainability. *Environment, Development and Sustainability, 26*, 3767–3792. https://doi.org/10.1007/s10668-022-02857-8

Manna, H., Mallick, S. K., Sarkar, S., & Roy, S. K. (2025). Developing decision making framework on built-up site suitability assessment for urban regeneration in the industrial cities of Eastern India. *Scientific Reports, 15*, 5708. https://doi.org/10.1038/s41598-025-90408-2

Martins, S., Bernardo, N., Ogashawara, I., & Alcantara, E. (2016). Support vector machine algorithm optimal parameterization for change detection mapping in Funil Hydroelectric Reservoir (Rio de Janeiro State, Brazil). *Modeling Earth Systems and Environment, 2*, 138. https://doi.org/10.1007/s40808-016-0190-y

Mohamed, A., & Worku, H. (2020). Simulating urban land use and cover dynamics using cellular automata and Markov chain approach in Addis Ababa and the surrounding. *Urban Climate, 31*, Article 100545. https://doi.org/10.1016/j.uclim.2019.100545

Mohammed, I., Alshuwaikhat, H. M., & Adenle, Y. A. (2016). An approach to assess the effectiveness of smart growth in achieving sustainable development. *Sustainability, 8*, 397. https://doi.org/10.3390/su8040397

Moreira, L. L., de Brito, M. M., & Kobiyama, M. (2021). Effects of different normalization, aggregation, and classification methods on the construction of flood vulnerability indexes. *Water, 13*, 98. https://doi.org/10.3390/w13010098

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*, 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001

Nechyba, T. J., & Walsh, R. P. (2002). Urban sprawl. *The Journal of Economic Perspectives, 18*, 177–200. https://doi.org/10.1257/0895330042632681

Ngu, N. H., Trung, N. H., Shinjo, H., Chotpantarat, S., & Thanh, N. N. (2025). Improving spatial prediction of soil organic matter in central Vietnam using Bayesian-enhanced machine learning and environmental covariates. *Archives of Agronomy and Soil Science, 71*, 1–17. https://doi.org/10.1080/03650340.2024.2448623

Parry, J. A., Ganaie, S. A., & Sultan Bhat, M. (2018). GIS based land suitability analysis using AHP model for urban services planning in Srinagar and Jammu urban centers of J&K, India. *J Urban Manag, 7*, 46–56. https://doi.org/10.1016/j.jum.2018.05.002

Peerzado, M. B., Magsi, H., & Sheikh, M. J. (2019). Land use conflicts and urban sprawl: Conversion of agriculture lands into urbanization in Hyderabad, Pakistan. *Journal of the Saudi Society of Agricultural Sciences, 18*, 423–428. https://doi.org/10.1016/j.jssas.2018.02.002

Phu, T. N. T., Thi, K. T., & Ky, P. N. (2023). *Assessment of the influence of urban flood in Thu Duc City in the period of planning* (pp. 62–76).

Phuong, D. N. D., Linh, V. T., Nhat, T. T., Dung, H. M., & Loi, N. K. (2019). Spatiotemporal variability of annual and seasonal rainfall time series in Ho Chi Minh city, Vietnam. *Journal of Water and Climate Change, 10*, 658–670. https://doi.org/10.2166/wcc.2018.115

Pinheiro, J. M. H., de Oliveira, S. V. B., Silva, T. H. S., Saraiva, P. A. R., de Souza, E. F., Godoy, R. V., Ambrosio, L. A., & Becker, M. (2025). The impact of feature scaling in machine learning: Effects on regression and classification tasks. *IEEE Access, 13*, 199903–199931. https://doi.org/10.1109/ACCESS.2025.3635541

Pourali, S. H., Arrowsmith, C., Chrisman, N., Matkan, A. A., & Mitchell, D. (2016). Topography wetness index application in flood-risk-based land use planning. *Applied Spatial Analysis and Policy, 9*, 39–54. https://doi.org/10.1007/s12061-014-9130-2

Qi, T., Li, Y., Huang, M., Luo, B., Peng, Z., Wang, W., Li, S., Degen, A. A., & Shang, Z. (2024). Land intensification use scenarios based on urban land suitability assessment of the national park. *Sustainable Cities and Society, 102*, Article 105229. https://doi.org/10.1016/j.scs.2024.105229

Quang, C. N. X., Giang, N. N. H., & Thinh, N. T. P. (2023). *Impacts of urbanization and climate change on surface runoff in Thu Duc City, Vietnam, 1170*, Article 012011. https://doi.org/10.1088/1755-1315/1170/1/012011

Ray, S. K., Teja, A., Chaudhary, S., Kumar, M. H., Rahul, R., Rammohan, K., & Himabindu, D. (2025). A framework for sustainable urban settlement planning: Integrating geographic information system and multi-criteria decision analysis: SK Ray et al. *Environment, Development and Sustainability*, 1–45. https://doi.org/10.1007/s10668-025-06442-7

Rimal, B., Zhang, L., Keshtkar, H., Haack, B. N., Rijal, S., & Zhang, P. (2018). Land use/land cover dynamics and modeling of urban land expansion by the integration of cellular automata and markov chain. *ISPRS International Journal of Geo-Information, 7*, 154. https://doi.org/10.3390/ijgi7040154

Rudra, S., Das, P., Mallick, S. K., Maity, B., Sahu, N., & Khan, U. (2025). Application of geo-spatial and geo-statistical techniques in site suitability analysis for potential urban built-up area development. *Environment, Development and Sustainability*, 1–29. https://doi.org/10.1007/s10668-024-05928-0

Sadian, A., & Shafizadeh-Moghadam, H. (2025). The future of agricultural lands under the combined influence of shared socioeconomic pathways and urban expansion by 2050. *Agricultural Systems, 224*, Article 104234. https://doi.org/10.1016/j.agsy.2024.104234

Saha, A., & Roy, R. (2021). An integrated approach to identify suitable areas for built-up development using GIS-based multi-criteria analysis and AHP in Siliguri planning area, India. *SN Applied Sciences, 3*, 395. https://doi.org/10.1007/s42452-021-04354-5

Saleem, J., Ahmad, S. S., & Shabbir, R. (2022). Evaluating land suitability analysis for urban services planning in coal clusters of Punjab using AHP, WOM, and TOPSIS method. *Arabian Journal of Geosciences, 15*, 324. https://doi.org/10.1007/s12517-022-09510-8

Saxena, V. (2025). Water quality, air pollution, and climate change: Investigating the environmental impacts of industrialization and urbanization. *Water, Air, & Soil Pollution, 236*, 73. https://doi.org/10.1007/s11270-024-07702-4

Saxena, A., & Jat, M. K. (2020). Land suitability and urban growth modeling: Development of SLEUTH-Suitability. *Computers, Environment and Urban Systems, 81*, Article 101475. https://doi.org/10.1016/j.compenvurbsys.2020.101475

Scott, A. J., Carter, C., Reed, M. R., Larkham, P., Adams, D., Morton, N., Waters, R., Collier, D., Crean, C., & Curzon, R. (2013). Disintegrated development at the rural–urban fringe: Re-connecting spatial planning theory and practice. *Progress in Planning, 83*, 1–52. https://doi.org/10.1016/j.progress.2012.09.001

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13*, 6308–6325. https://doi.org/10.1109/JSTARS.2020.3026724

Shikhteymour, S. R., Borji, M., Bagheri-Gavkosh, M., Azimi, E., & Collins, T. W. (2023). A novel approach for assessing flood risk with machine learning and multi-criteria decision-making methods. *Applied Geography, 158*, Article 103035. https://doi.org/10.1016/j.apgeog.2023.103035

Sihasale, D. A., Latue, P. C., & Rakuasa, H. (2023). Spatial analysis of built-up land suitability in ternate island. *Jurnal Riset Multidisiplin Dan Inovasi Teknologi, 1*, 70–83. https://doi.org/10.59653/jimat.v1i02.219

Singh, L., Saravanan, S., Jennifer, J. J., & Abijith, D. (2021). Application of multi-influence factor (MIF) technique for the identification of suitable sites for urban settlement in Tiruchirappalli City, Tamil Nadu, India. *Asia-Pacific Journal of Regional Science, 5*, 797–823. https://doi.org/10.1007/s41685-021-00194-8

Sun, L., Tang, L., Shao, G., Qiu, Q., Lan, T., & Shao, J. (2019). A machine learning-based classification system for urban built-up areas using multiple classifiers and data sources. *Remote Sensing, 12*, 91. https://doi.org/10.3390/rs12010091

Taalab, K., Cheng, T., & Zhang, Y. (2018). Mapping landslide susceptibility and types using Random Forest. *Big Earth Data, 2*, 159–178. https://doi.org/10.1080/20964471.2018.1472392

Thanh, N. N., Chotpantarat, S., Trung, N. H., Ngu, N. H., & Van Muoi, L. (2022). Mapping groundwater potential zones in Kanchanaburi Province, Thailand by integrating of analytic hierarchy process, frequency ratio, and random forest. *Ecological Indicators, 145*, Article 109591. https://doi.org/10.1016/j.ecolind.2022.109591

Thapa, N., Pant, P., Prasai, R., Mahata, A., & Dulal, S. (2025). Sustainable land use planning in developing countries using GIS and multi-criteria analysis: A case study of Lalitpur district, Nepal. *City and Built Environment, 3*, 13. https://doi.org/10.1007/s44213-025-00050-x

Thi, T. N. K., Phung, H. P., Thy, P. T. M., Nguyen, L. D., & Thanh, N. H. (2025). In *Monitoring and analysis of hot spots of land surface temperature to understand the surface urban heat island effect in Thu Duc City*, Article 012012. https://doi.org/10.1088/1755-1315/1501/1/012012, 2025.

Tuan, N. T., Hegedűs, G., & Phuong, N. T. T. (2025). Urbanization and forecast possibilities of land use changes by 2050: New evidence in Ho Chi Minh city, Vietnam. *The Open Agriculture, 10*, Article 20250421. https://doi.org/10.1515/opag-2025-0421

Ullah, K. M., & Mansourian, A. (2016). Evaluation of land suitability for urban land-use planning: Case study D haka City. *Transactions in GIS, 20*, 20–37. https://doi.org/10.1111/tgis.12137

Utami, N. D. N., Supriatna, S., & Anggrahita, H. (2018). Multi criteria analysis of Built-Up land suitability in sleman regency, Special Region of Yogyakarta Province, 2018 . https://doi.org/10.1051/e3sconf/20187303005.

Virtriana, R., Ihsan, K. T. N., Anggraini, T. S., Deliar, A., Harto, A. B., Riqqi, A., & Sakti, A. D. (2025). Development of location suitability prediction for health facilities using random forest machine learning in 2030 integrating remote sensing and GIS in West Java, Indonesia. *Environmental Advances, 19*, Article 100604. https://doi.org/10.1016/j.envadv.2024.100604

Weldu, W. G., & Deribew, I. A. (2016). Identification of potential sites for housing development using GIS based multi-criteria evaluation in Dire Dawa City, Ethiopia. *International Journal of Sciences: Basic and Applied Research, 28*, 34–49.

Wu, B., Wang, S., Zou, Y., Chen, Y., Luo, Y., Chen, Y., & Li, Y. (2025). Multi-scenario simulation of land use spatial conflicts: A spatially explicit model integrating conflict risk and ecological impact: B. Wu et al. *Environment, Development and Sustainability*, 1–28. https://doi.org/10.1007/s10668-025-06329-7

Wu, R., Wang, J., Zhang, D., & Wang, S. (2021). Identifying different types of urban land use dynamics using point-of-interest (POI) and Random Forest algorithm: The case of Huizhou, China. *Cities, 114*, Article 103202. https://doi.org/10.1016/j.cities.2021.103202

Wu, K.-y., & Zhang, H. (2012). Land use dynamics, built-up land expansion patterns, and driving forces analysis of the fast-growing Hangzhou metropolitan area, eastern China (1978–2008). *Applied Geography, 34*, 137–145. https://doi.org/10.1016/j.apgeog.2011.11.006

Xi, C.-b., Qian, T.-l., Chi, Y., Chen, J., & Wang, J.-c. (2018). Relationship between settlements and topographical factors: An example from Sichuan Province, China. *Journal of Mountain Science, 15*, 2043–2054. https://doi.org/10.1007/s11629-018-4863-z

Zhou, X., Zhang, S., Liu, Y., Zhou, Q., Wu, B., Gao, Y., & Zhang, T. (2022). Impact of urban morphology on the microclimatic regulation of water bodies on waterfront in summer: A case study of Wuhan. *Building and Environment, 226*, Article 109720. https://doi.org/10.1016/j.buildenv.2022.109720