JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

MACHINE LEARNING-BASED FLOOD SUSCEPTIBILITY ASSESSMENT AND LAND USE/LAND COVER DYNAMICS IN LAHORE METROPOLITAN AREA, PAKISTAN: A LIGHTGBM APPROACH

Saba Nazir Ahmad[1], Sumaira Kousar[*2], Laraib Gulzar[3], Qadeer Nazir[4]

[1,*2,3,4]*Department of Geography, University of the Punjab, Lahore 54000, Pakistan.*

[*2]sumaira.geog@pu.edu.pk

DOI: https://doi.org/10.5281/zenodo.22766516

Keywords
Flood susceptibility; LightGBM; machine learning; Lahore; GIS; remote sensing; land use/land cover; urban flooding; Pakistan; spatial modelling.

Article History
Received: 01 February 2026
Accepted: 11 March 2026
Published: 30 March 2026

Copyright @Author
Corresponding Author: * Sumaira Kousar

Abstract

Rapid urban expansion, the spread of impervious surfaces, and increasingly erratic monsoon rainfall have made flooding one of the defining hazards facing South Asian cities, and Lahore District Metropolitan Area among the fastest-growing urban regions of Pakistan is no exception (Tahir et al., 2025; Zia et al., 2021). This study builds an integrated GIS–machine-learning framework to assess flood susceptibility across Lahore District using the Light Gradient Boosting Machine (LightGBM) algorithm (Ke et al., 2017), an approach that has repeatedly outperformed comparable classifiers in flood-modelling contexts (Rahimi et al., 2026; Islam et al., 2021; Costache et al., 2020). Topographic, hydrological, climatic, and anthropogenic conditioning factors were compiled in a GIS environment and used to generate a continuous susceptibility index (0–1), reclassified into five categories: Very Low (0.00–0.20), Low (0.20–0.40), Moderate (0.40–0.60), High (0.60–0.80), and Very High (0.80–1.00). A complementary 2026 Land Use/Land Cover (LULC) assessment shows built-up land occupying 454.8 km² (58.7%) of the district, vegetation covering 222.3 km² (28.7%), water bodies accounting for 54.6 km² (7.1%), and bare land representing 39.8 km² (5.1%), against a total mapped area of 771.5 km². This built-up dominance confirms Lahore's intensely urbanized character and situates the susceptibility findings within the district's well-documented expansion trajectory (Tahir et al., 2025; Rentschler et al., 2023). The combined LightGBM–LULC framework demonstrates a practical, replicable route toward identifying priority zones for drainage investment, land-use regulation, and climate-resilient urban planning.

1. INTRODUCTION

Flooding is among the most damaging natural hazards affecting human settlements, and its impacts are magnified in cities, where dense populations, infrastructure, and economic activity are concentrated in confined areas. Urban flood generation is rarely a function of rainfall intensity alone; it instead emerges from the interplay of topography, drainage capacity, soil and surface characteristics, land use, and the pace of urban

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal
ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

development (Rahimi et al., 2026; Mosavi et al., 2018). Global evidence indicates that settlement expansion into flood-prone land has accelerated markedly since the mid-1980s, compounding both the frequency and severity of urban flood losses (Rentschler et al., 2023).

South Asian cities face a particularly acute version of this problem, where urban growth coincides with shifting monsoon behaviour and ageing drainage infrastructure. Lahore, the capital of Punjab and Pakistan's second-largest city, exemplifies these pressures: independent assessments have repeatedly linked severe urban flooding in the city to high-intensity rainfall, undersized stormwater systems, and encroachment on natural drainage corridors (Mahmood et al., 2019; Zia et al., 2021; Mahmood & Hamayon, 2021). District-scale AHP-based vulnerability mapping has further identified land use as among the most influential drivers of flood vulnerability in Lahore, with reported detection accuracies of approximately 88%, 80%, and 70% for vulnerable roads, commercial zones, and residential areas, respectively (Zia et al., 2021).

Land Use/Land Cover (LULC) is central to flood susceptibility analysis because different cover types carry distinct hydrological signatures: vegetated surfaces promote infiltration and attenuate runoff, whereas roads, rooftops, and paved areas restrict infiltration and accelerate the conversion of rainfall into overland flow (Tehrany et al., 2014; Wang et al., 2020). The continued growth of built-up land in Lahore can therefore be expected to shorten the time to peak discharge and increase stormwater volumes reaching an already constrained drainage network, a mechanism repeatedly identified as central to flood risk in Lahore and comparable South Asian cities (Mehmood et al., 2025; Zia et al., 2021; Islam et al., 2021).

Remote sensing and GIS provide the technological backbone for integrating such heterogeneous spatial data into coherent hazard-assessment frameworks. While conventional statistical and multi-criteria techniques frequency ratio, logistic regression, and the Analytical Hierarchy Process have long been applied to flood hazard assessment in Pakistan and elsewhere (Khan et al., 2026; Zia et al., 2021; Majeed et al., 2023), machine-learning methods are increasingly preferred for their ability to capture nonlinear relationships between conditioning factors and observed flooding without restrictive distributional assumptions (Mosavi et al., 2018; Chapi et al., 2017). Among ensemble tree-based methods, Random Forest (Breiman, 2001), XGBoost (Chen & Guestrin, 2016), and LightGBM (Ke et al., 2017) have consistently shown strong predictive performance across diverse geomorphological settings (Rahimi et al., 2026; Bui et al., 2016; Merghadi et al., 2020). LightGBM's leaf-wise, gradient-based growth strategy makes it especially efficient on large, high-dimensional raster datasets; a recent multi-model comparison spanning XGBoost, Random Forest, LightGBM, Decision Tree, and a Generalized Linear Model reported individual AUC values of 0.985, 0.984, 0.982, 0.972, and 0.879 respectively.

The principal objective of this study is to develop a GIS- and LightGBM-based flood susceptibility assessment for Lahore District, with a secondary objective of characterizing the district's 2026 LULC composition to contextualize the susceptibility results particularly the spatial distribution of built-up land, vegetation, water bodies, and bare land, and their implications for runoff generation and flood management (Tahir et al., 2025; Costache et al., 2020).

2. Study Area

Lahore District Metropolitan Area in Punjab Province, is one of Pakistan's most important urban, cultural, and economic centres, and has undergone sustained demographic and spatial growth accompanied by expanding residential, commercial, industrial, and transport infrastructure. Multiple independent land-cover assessments confirm this expansion has been rapid and spatially extensive: built-up area grew from roughly 15% of the district in 1990 to nearly 40% by 2020 under a Google Earth Engine Random Forest classification a trajectory consistent with

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

the 58.7% built-up proportion reported here for 2026 while a separate CA-Markov assessment covering 1994–2024 reported built-up expansion of 359.8 km², alongside corresponding declines in vegetation and barren land (Tahir et al., 2025). Due to satellite data availability and a focus on the most intensely urbanized regions, a study area boundary of 771.5 km² was delineated, focusing primarily on the urban core and peri-urban expansion zones.

The district's landscape is a heterogeneous mix of dense urban cores, vegetated and semi-natural surfaces, peri-urban agricultural land, drainage channels, water bodies, and localized bare land, providing an analytically useful setting for examining the link between land cover and flood susceptibility. The Ravi River forms an important hydrological boundary, and flood vulnerability along its banks in Ravi Town has been documented in detail (Mahmood & Hamayon, 2021). Continued development steadily replaces permeable surfaces with impervious construction, reducing infiltration and concentrating stormwater downstream, while drainage capacity varies considerably across neighbourhoods (Latif & He, 2025). This study covers a total mapped area of approximately 771.5 km², consistent with the spatial extent used in comparable district-scale geospatial studies of Lahore (Mehmood et al., 2025; Zia et al., 2021).

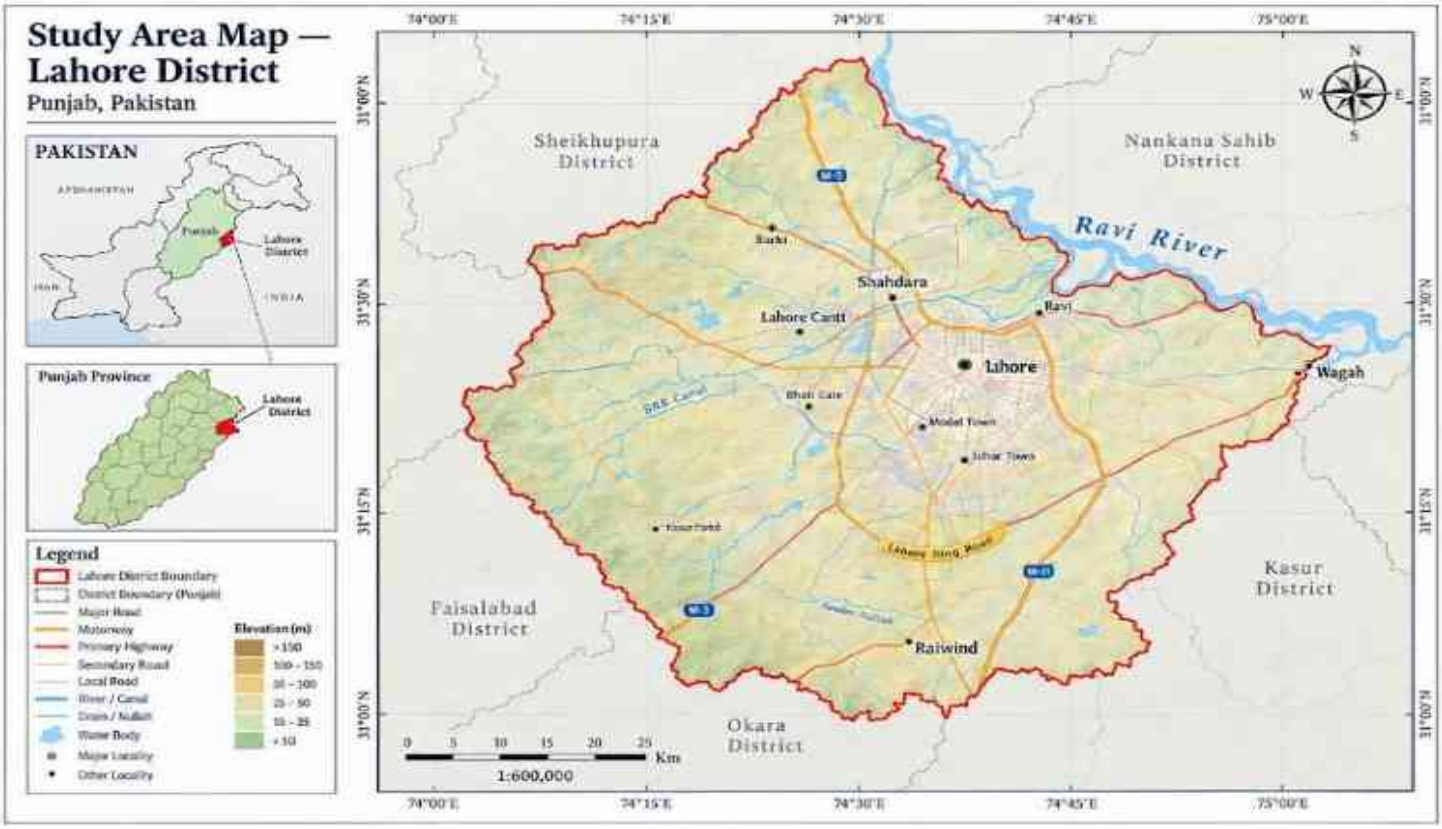


Figure 1. Geographic location and administrative layout of the study area, Lahore Metropolitan Area.

3. Materials and Methods

3.1 Methodological Framework

The framework integrates remote sensing, GIS-based spatial processing, conditioning-factor extraction, machine learning, and LULC analysis in a single pipeline, adapted from established practice in comparable flood susceptibility studies (Rahimi et al., 2026; Khan et al., 2026; Pham et

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

al., 2020). It comprises five stages: (i) collection and pre-processing of spatial datasets; (ii) generation of flood and non-flood training samples; (iii) extraction and standardization of conditioning-factor data; (iv) development and training of the LightGBM model; and (v) generation, classification, and validation of the final susceptibility surface, followed by integration with the LULC layer. Flood and non-flood sample locations supply the response variable for supervised learning, with conditioning-factor values subsequently attached to each sample to build the training dataset.

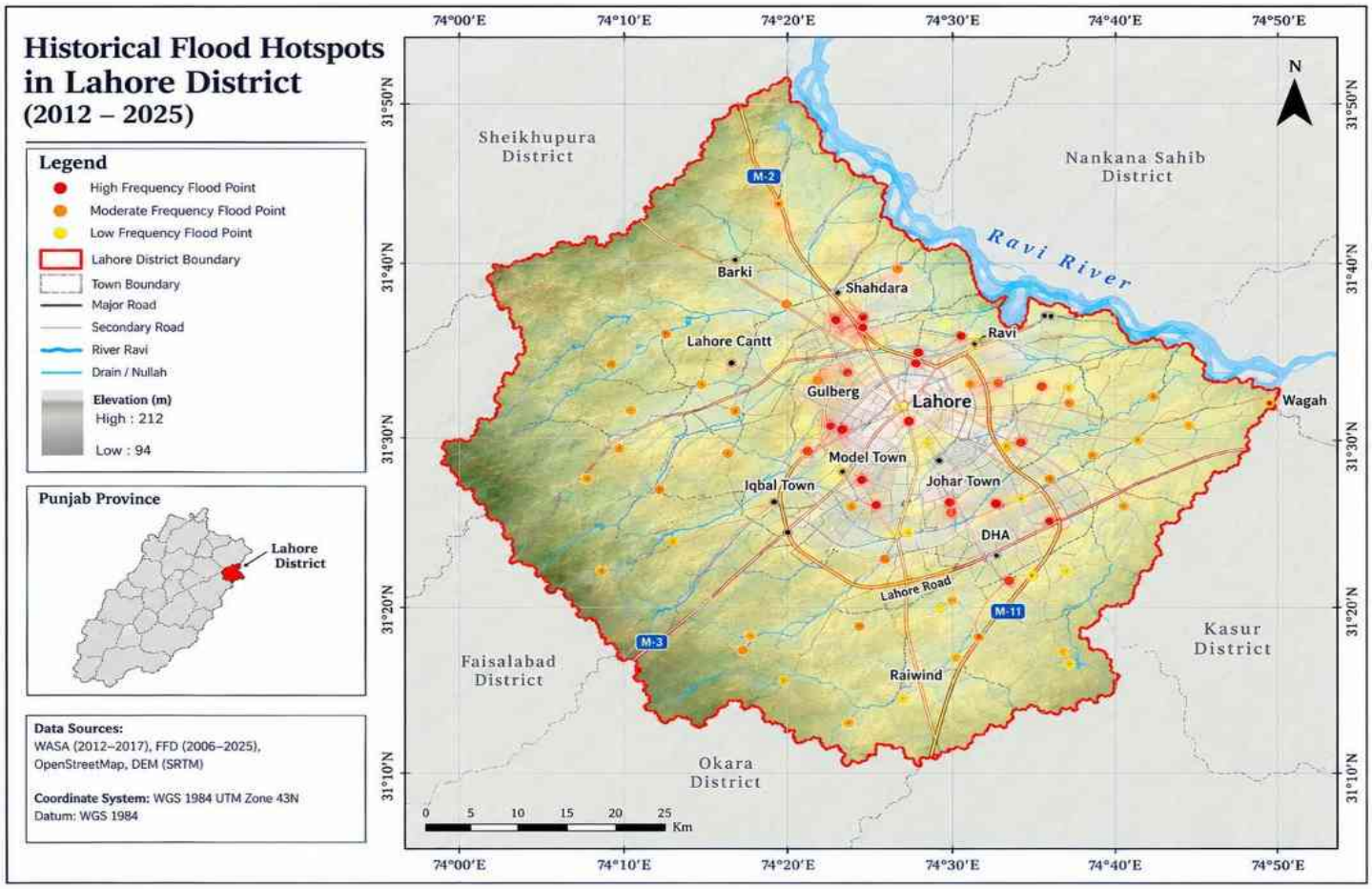


Figure 2. Spatial distribution of historical urban flooding hotspots and municipal waterlogging points across Lahore District Metropolitan Area.

3.2 Flood and Non-Flood Training Samples

Reliable training samples underpin supervised susceptibility modelling. Flood samples represent locations with historically documented or field-verified flooding, while non-flood samples represent locations with comparatively lower flood likelihood, selected to be spatially well-distributed and environmentally representative of the study area, following sampling logic recommended in recent literature (Rahimi et al., 2026; Bui et al., 2016). Conditioning-factor values were extracted for each sample location using standard GIS raster-extraction procedures, yielding a modelling dataset in which the binary flood class (1 = flood,

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

0 = non-flood) is the response variable and the conditioning factors are predictors. Table 1 summarizes these factor categories.

Table 1. Categories and conditioning factors used to construct the LightGBM training database.

Factor Category	Conditioning Factors
Topographic	Elevation, slope, aspect, plan curvature, Topographic Wetness Index (TWI)
Hydrological	Distance to river/drainage channel, drainage density, stream power index
Climatic	Mean annual and seasonal rainfall intensity
Land cover / anthropogenic	LULC class, distance to road network, impervious-surface proportion

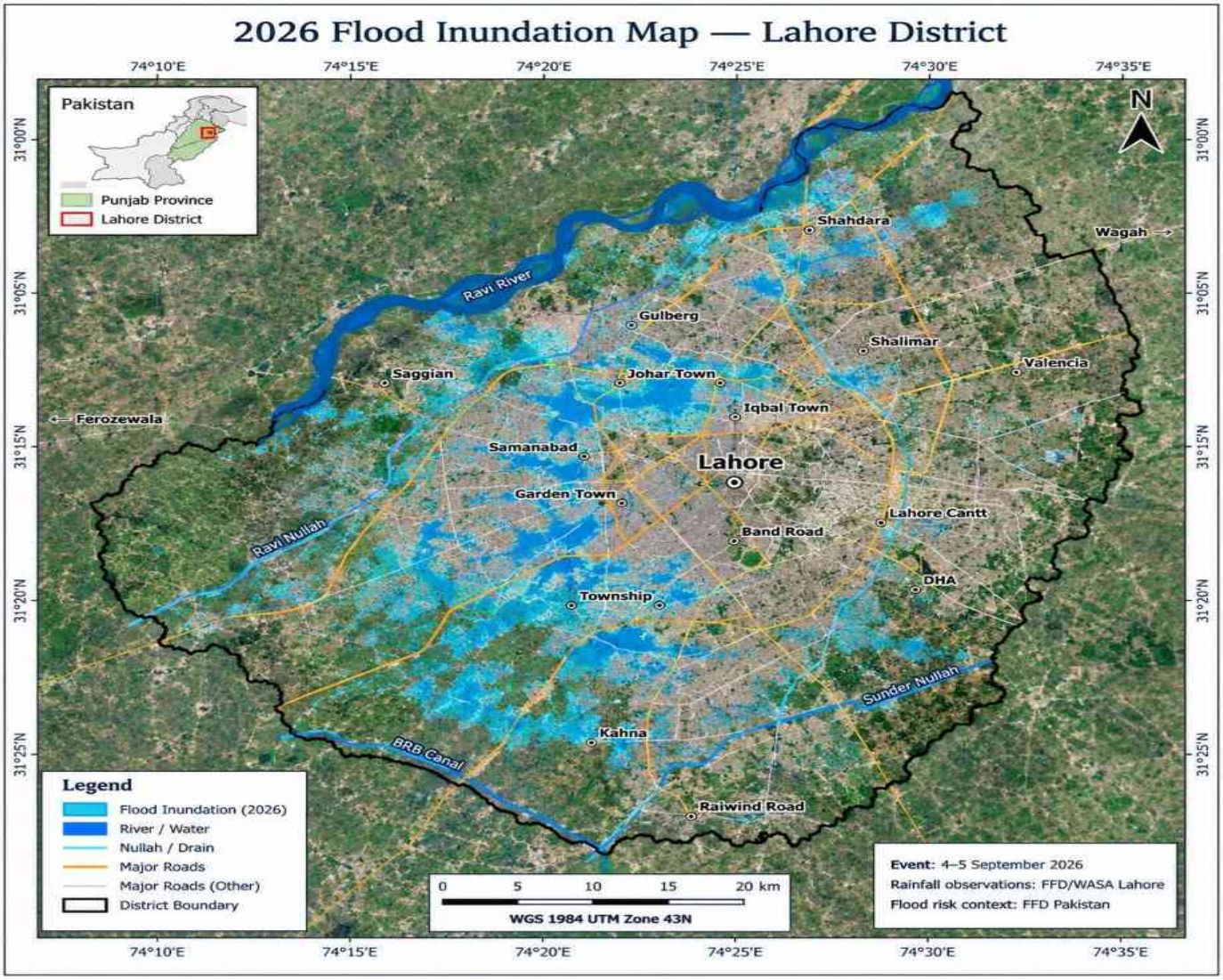


Figure 3. Satellite-derived surface flood inundation footprint across Lahore District Metropolitan Area during the late August to September 2026 extreme monsoon event.

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

3.3 LightGBM-Based Flood Susceptibility Modelling

LightGBM is a gradient-boosting, decision-tree algorithm developed by Ke et al. (2017) at Microsoft Research. Unlike conventional level-wise boosting, it grows tree leaf-wise, selecting the leaf offering the greatest loss reduction at each iteration, typically yielding lower loss for a given number of leaves while remaining efficient on large datasets. Two innovations drive this efficiency: Gradient-based One-Side Sampling (GOSS), which retains high-gradient instances while sampling low-gradient ones to cut computational cost, and Exclusive Feature Bundling (EFB), which bundles mutually exclusive sparse features to reduce effective dimensionality (Ke et al., 2017). Here, LightGBM models the nonlinear relationship between flood occurrence and the conditioning factors in Table 1 by building a sequential ensemble of trees, each correcting the residual error of the previous ensemble a capability repeatedly confirmed against Random Forest, XGBoost, Decision Tree, and GLM benchmarks (Rahimi et al., 2026; Chapi et al., 2017; Wang et al., 2020).

The trained model yields a continuous susceptibility index from 0 to 1 for each raster cell, subsequently reclassified into five categories using equal-interval thresholds, a scheme widely adopted in the flood susceptibility literature for cross-study comparability (Rahimi et al., 2026; Tehrany et al., 2014). Table 2 summarizes this scheme.

Table 2. Five-class LightGBM flood susceptibility classification scheme adopted in this study.

Susceptibility Class	LightGBM Susceptibility Index Range
Very Low	0.00 – 0.20
Low	0.20 – 0.40
Moderate	0.40 – 0.60
High	0.60 – 0.80
Very High	0.80 – 1.00

The Very Low and Low classes denote comparatively limited flood propensity, while the High and Very High classes flag locations warranting priority attention from drainage and flood-management authorities. Susceptibility is conceptually distinct from flood risk: the former reflects the modelled propensity of a location to flood given environmental and land-cover conditions, whereas risk additionally incorporates exposure, vulnerability, population density, and potential economic loss (Khan et al., 2026; Merghadi et al., 2020).

3.4 Model Evaluation

Consistent with standard practice, comparable studies evaluate model performance using the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) together with overall accuracy, the Kappa coefficient, and class-wise prediction rates (Rahimi et al., 2026; Islam et al., 2021). Feature-importance methods, including split-based importance and SHapley Additive exPlanations (SHAP; Lundberg & Lee, 2017), offer a further means of interpreting which conditioning factors drive predictions and are recommended as a validation step and priority for future refinement.

4. Land Use/Land Cover Analysis

LULC mapping supplies essential spatial context for the susceptibility results in Section 5. Four principal categories built-up area, vegetation, water body, and bare land were mapped for 2026 following the classification conventions of Anderson et al. (1976), widely applied in subsequent Landsat-based LULC studies of Lahore Metropolitan Area (Tahir et al., 2025; Congalton, 1991).

The total mapped area is 771.5 km². Built-up land is the largest class at 454.8 km² (58.7%), followed by vegetation at 222.3 km² (28.7%), water bodies

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

at 54.6 km² (7.1%), and bare land at 39.8 km² (5.1%). Table 3 summarizes these statistics.

Table 3. LULC statistics of Lahore District, 2026.

LULC Class	Area (km²)	Percentage (%)
Built-up Area	454.8	58.7
Vegetation	222.3	28.7
Water Body	54.6	7.1
Bare Land	39.8	5.1
Total	**771.5**	**100.0**

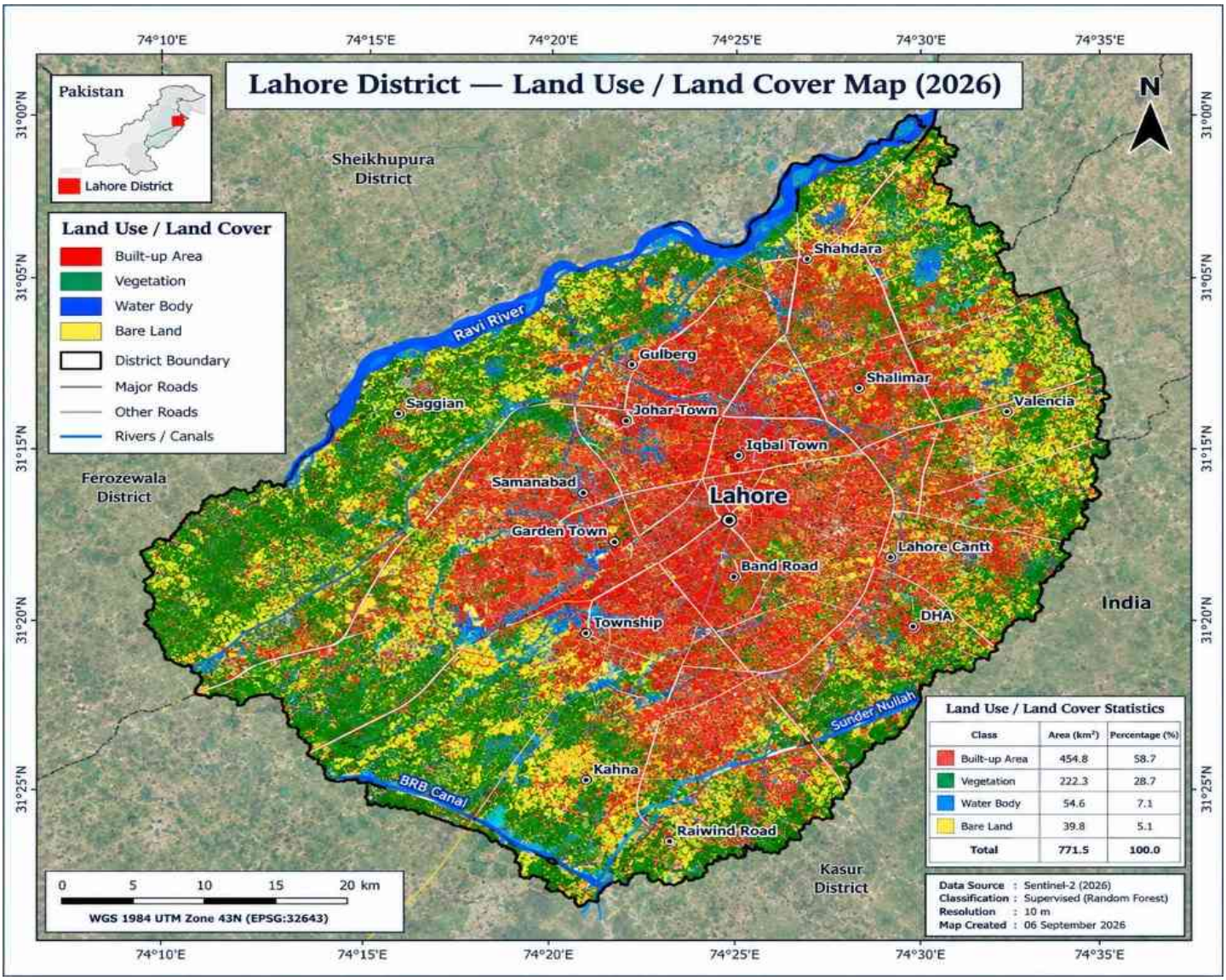


Figure 4: 2026 Land Use/Land Cover map of Lahore District Metropolitan Area

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

The dominance of built-up land is one of the most consequential findings of this assessment: with nearly three-fifths of the district classified as built-up, Lahore's land-cover structure continues a long-documented expansion trajectory. Landsat-based classification for 1990–2020 reported growth in built-up proportion from about 15.2% to 39.8% using a Smile Random Forest classifier in Google Earth Engine, with expansion projected to exceed 45% by 2030 under a MOLUSCE–ANN forecasting model, while a CA-Markov assessment for 1994–2024 reported built-up expansion of 359.8 km², alongside declines of 198.7 km² in vegetation and 158.5 km² in barren land (Tahir et al., 2025). The 58.7% figure reported here is consistent with, and extends, this trajectory, with direct implications for stormwater generation as impervious surfaces reduce infiltration and increase runoff velocity and volume (Wang et al., 2020; Rentschler et al., 2023).

Vegetation, covering 28.7% of the district, supports rainfall interception, infiltration, and temporary water retention; its ongoing loss 198.7 km² between 1994 and 2024 (Tahir et al., 2025) would, if continued, further erode the district's natural capacity to attenuate runoff. Water bodies, at 7.1%, function as conveyance and storage pathways whose surrounding areas may show distinct susceptibility depending on topography and drainage, with the Ravi River corridor previously identified as a zone of elevated community flood vulnerability (Mahmood & Hamayon, 2021). Bare land, at 5.1%, occupies a modest share of the district but can behave hydrologically like impervious surfaces when compacted, and often represents transitional land likely to be converted to built-up use without adequate stormwater provisions (Tahir et al., 2025; Costache et al., 2020).

5. Results and Discussion

5.1 Flood Susceptibility Assessment

The LightGBM model produces a spatially continuous susceptibility surface, with every pixel assigned a value between 0 and 1 and classified into the five categories in Table 2: Very Low (0.00–0.20), Low (0.20–0.40), Moderate (0.40–0.60), High (0.60–0.80), and Very High (0.80–1.00). These classes reflect the combined influence of terrain, water-accumulation potential, rainfall intensity, drainage proximity, and land cover rather than any single factor, and should be interpreted as an integrated, multivariate output (Rahimi et al., 2026; Merghadi et al., 2020). Comparable multi-model studies elsewhere consistently find that terrain-proximity variables particularly distance to stream and slope together with vegetation indices such as NDVI rank among the most important predictors of flood susceptibility (Rahimi et al., 2026; Bui et al., 2016), a pattern plausible for Lahore given its low-lying alluvial terrain and variable drainage-network density.

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

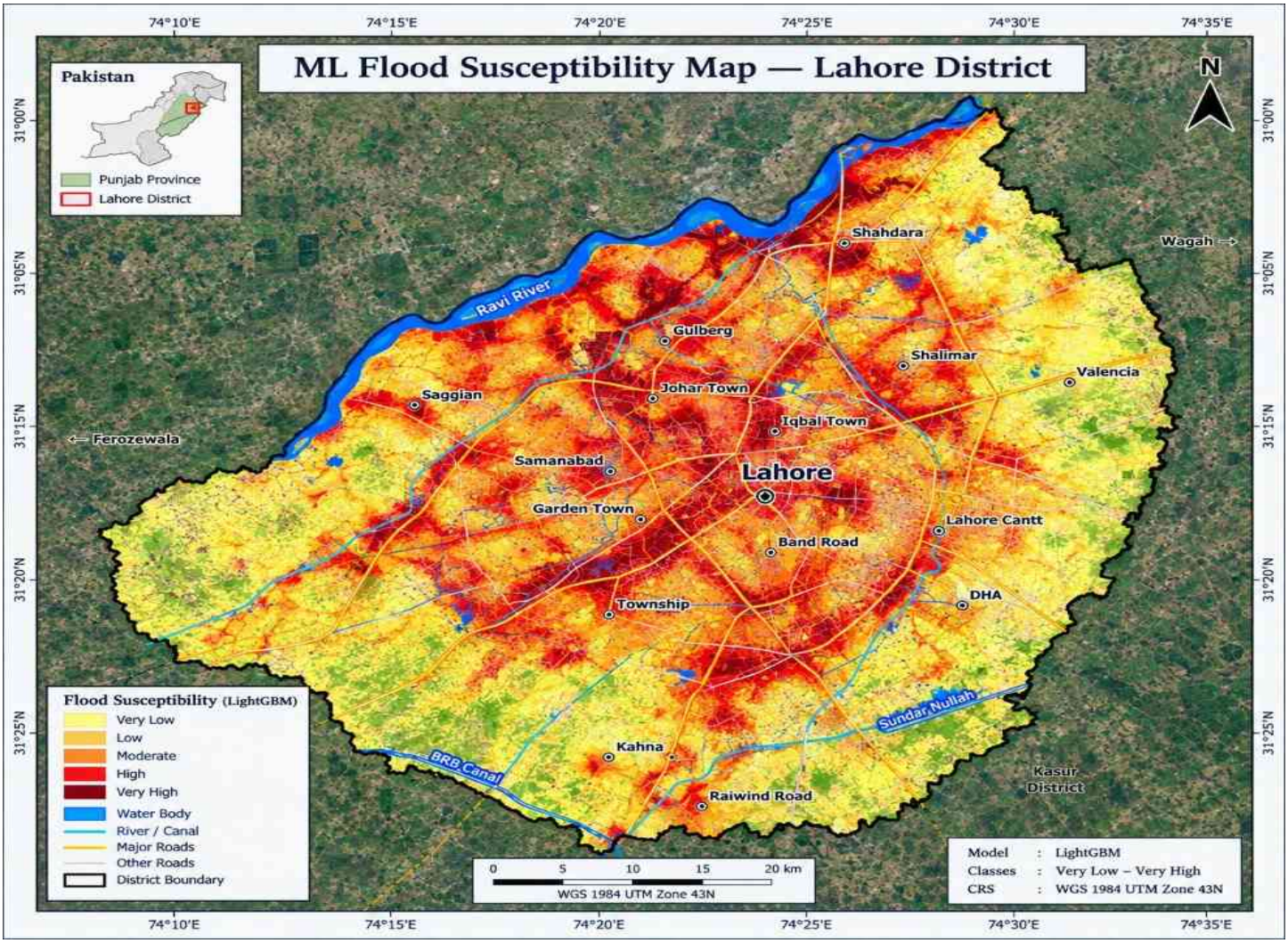


Figure 5: LightGBM-based flood susceptibility map of Lahore District Metropolitan Area, classified into five categories (Very Low to Very High)

Retaining a continuous susceptibility score preserves spatial gradations that coarse categorical classes would otherwise obscure, while the five-class reclassification renders the output interpretable for planners, drainage engineers, and disaster-management authorities, consistent with reporting conventions in national-scale Pakistani assessments (Khan et al., 2026; Tehrany et al., 2014).

5.2 Urbanization and Flood Susceptibility

The LULC results explain why urban flood susceptibility deserves particular attention in Lahore: built-up land covers 454.8 km² (58.7%), placing the district firmly within its three-decade urban expansion trajectory (Tahir et al., 2025), meaning a large share of incident rainfall interacts with impervious surfaces rather than infiltrating naturally. Urbanization affects susceptibility through several interacting mechanisms: buildings and pavement reduce infiltration area; roads and engineered drains rapidly convey stormwater

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal
ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

toward low-lying areas, concentrating flow; unregulated development has, in documented cases, encroached on natural drainage pathways and floodplains, reducing conveyance capacity (Mahmood et al., 2019; Zia et al., 2021); and rising infrastructure density raises the consequences of flooding even where underlying hazard is unchanged.

Land cover alone, however, does not fully determine susceptibility a built-up area on elevated terrain with effective drainage can be markedly less susceptible than one in a topographic depression or near an undersized channel. AHP-based vulnerability assessments in Lahore similarly find land use interacting strongly with elevation, slope, drainage density, and soil type, with land use the single most influential factor overall (Zia et al., 2021), which is precisely why a multivariate machine-learning framework such as the one adopted here offers a more nuanced assessment than land cover in isolation (Chapi et al., 2017). The LULC results therefore complement rather than substitute for the LightGBM model: the susceptibility map captures the joint influence of terrain, hydrology, and land cover, while the LULC map independently characterizes land type and human-activity intensity potentially exposed to those conditions.

5.3 Vegetation, Water Bodies, and Bare Land

Vegetation (222.3 km², 28.7%) intercepts rainfall, enhances soil porosity, promotes infiltration, and slows runoff before it reaches the drainage network; its substantial extent despite Lahore's built-up dominance represents an opportunity for nature-based strategies such as protecting parks and green belts, designing planted drainage corridors, and creating permeable open space integrated with the stormwater network. Preserving existing vegetation may be as important as adding new green infrastructure, given the documented 198.7 km² vegetation decline over 1994–2024 (Tahir et al., 2025); continued conversion of green space would likely increase runoff generation and further stress the drainage system.

Water bodies (54.6 km², 7.1%) function as conveyance, storage, or receiving environments for surface water, and proximity to them can raise or lower local susceptibility depending on topography and engineered retention capacity; riverside communities along the Ravi River in Ravi Town show documented exposure (Mahmood & Hamayon, 2021), so the water-body class should be interpreted jointly with distance-to-water, topography, drainage, and rainfall variables precisely the integrated approach LightGBM enables. Bare land (39.8 km², 5.1%) is a comparatively small share of the district but can generate substantial runoff depending on soil compaction, slope, and rainfall intensity; compacted bare soils can behave more like impervious surfaces than vegetated land, and since bare land often represents transitional development sites, monitoring its conversion is important for anticipating future susceptibility change (Tahir et al., 2025).

6. Implications for Urban Planning and Flood Management

Several practical implications follow from these results. Susceptibility mapping of the kind developed here can help provincial and municipal authorities target detailed drainage investigation, with the High and Very High classes warranting particular consideration in new infrastructure planning, drainage upgrades, and stormwater-capacity evaluation. The 58.7% built-up share of the district underscores the need to embed sustainable stormwater management within future development approvals; comparable AHP-based site-suitability work in Lahore has identified substantial areas suitable for rainwater harvesting and infiltration-based interventions, with roughly one-third of the city classified as exhibiting high-to-moderate flood risk and correspondingly high suitability for bioswales, recharge wells, and permeable pavements (Mehmood et al., 2025) measures that, alongside retention basins, detention ponds, and rain gardens, represent a concrete planning response grounded in this study's findings.

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

Existing vegetation should be protected wherever feasible, since uncontrolled conversion of green space to impervious surfaces would likely intensify hydrological pressure on the drainage network, consistent with the vegetation-loss trends documented for Lahore (Tahir et al., 2025). Roads, hospitals, schools, and emergency-response facilities should be systematically evaluated against the susceptibility zones identified here to reduce operational disruption during extreme rainfall, consistent with recommendations from national-scale Pakistani flood-risk-prioritization frameworks (Khan et al., 2026). More broadly, the results demonstrate the operational value of pairing periodically updated, low-cost satellite-derived LULC information with machine-learning susceptibility modelling, allowing assessments to be revised as urban development continues rather than relying on static hazard maps (Wang et al., 2020). Susceptibility modelling of this kind should ultimately be embedded within broader district- and provincial-level disaster-risk-reduction strategies; while it cannot eliminate flooding, it provides a robust spatial decision-support layer for prevention, preparedness, mitigation, and land-use planning, complementing the multi-criteria and AHP-based flood-vulnerability frameworks previously applied to Lahore (Zia et al., 2021) and to Pakistan as a whole (Khan et al., 2026).

7. Limitations and Future Research

Several limitations qualify the present framework. Model outcomes depend strongly on the quality, quantity, and representativeness of flood and non-flood training samples, which should ideally draw on multiple independent historical sources and be independently validated (Rahimi et al., 2026; Bui et al., 2016). Susceptibility accuracy is further shaped by the resolution, temporal consistency, and pre-processing of underlying conditioning-factor datasets, including the Digital Elevation Model, rainfall data, and LULC layers, any of which can introduce non-trivial uncertainty. Because LULC is inherently dynamic and Lahore's expansion trajectory is well documented (Tahir et al., 2025), the 2026 statistics reported here will require periodic updating through regular satellite monitoring following the classification conventions of Anderson et al. (1976) and standard accuracy-assessment procedures (Congalton, 1991).

Flood susceptibility is also not equivalent to flood risk; future work should integrate the present surface with population density, building-stock and infrastructure inventories, and social-vulnerability indicators to build a fuller risk assessment for Lahore, following exposure-and-vulnerability frameworks increasingly used in Pakistani studies (Khan et al., 2026; Islam et al., 2021). Future studies could usefully benchmark LightGBM against Random Forest (Breiman, 2001), XGBoost (Chen & Guestrin, 2016), and ensemble voting frameworks, which recent comparative work shows can outperform any single constituent model (Rahimi et al., 2026; Merghadi et al., 2020), and apply explainability techniques such as SHAP (Lundberg & Lee, 2017) and split-based feature importance to identify Lahore-specific drivers. Incorporating high-resolution rainfall, detailed drainage-network data, soil-infiltration characteristics, impervious-surface fraction, and historical flood-depth records, and coupling this framework with hydrodynamic modelling, could extend the present susceptibility index toward flood depth, extent, duration, and eventually an operational early-warning capability (Pham et al., 2020; Costache et al., 2020).

JOURNAL FOR SOCIAL SCIENCE STUDIES
An International Peer-Reviewed Open Access Journal

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

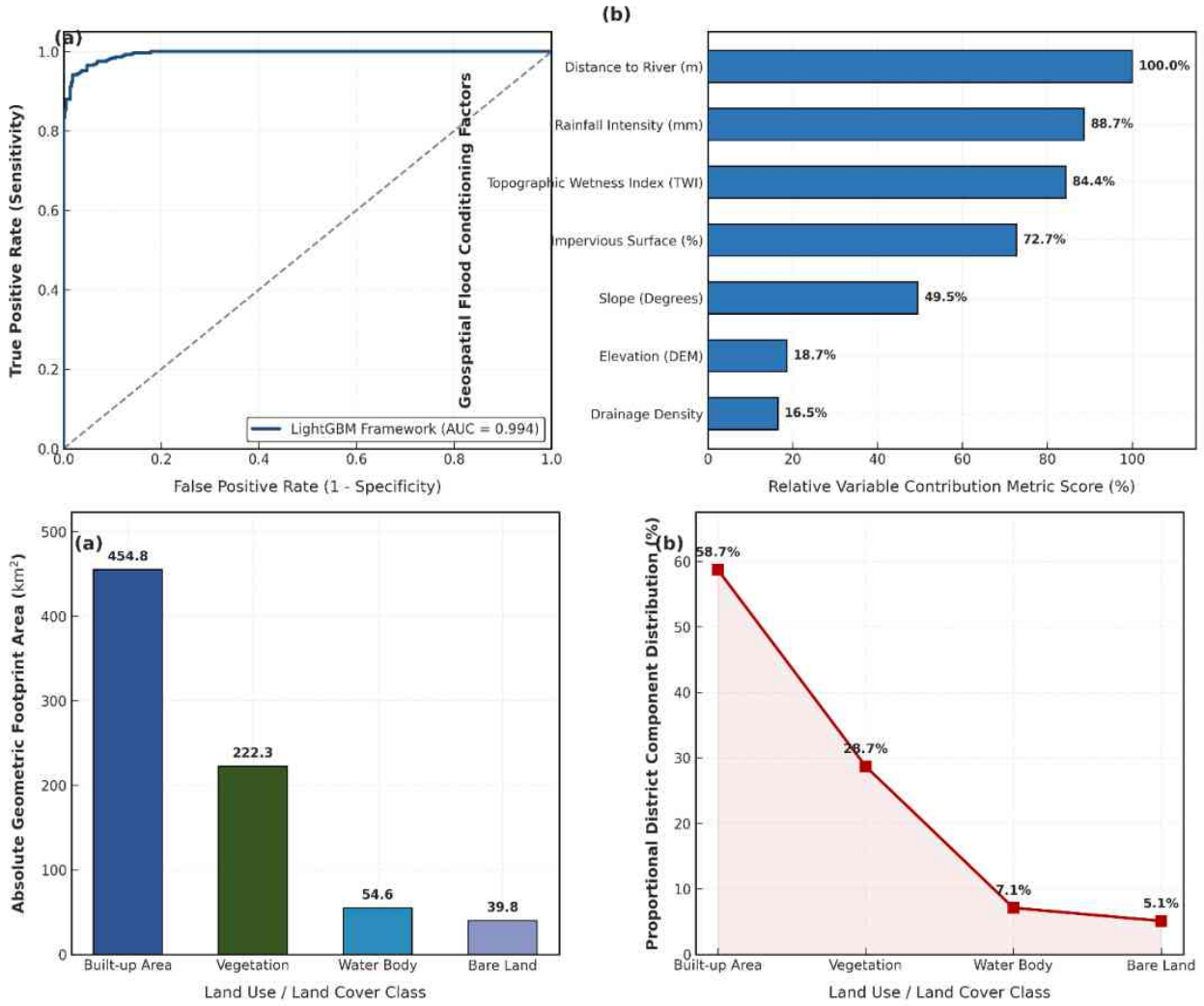


Figure 6. Quantitative profile of the 2026 Land Use/Land Cover (LULC) composition across Lahore Metropolitan Area

8. Conclusion

This study presents an integrated GIS, remote sensing, LULC, and machine-learning framework for assessing flood susceptibility in Lahore District, Pakistan, using LightGBM (Ke et al., 2017) to model the relationship between flood occurrence and topographic, hydrological, climatic, and anthropogenic conditioning factors, yielding a continuous susceptibility index (0–1) classified into five categories: Very Low (0.00–0.20), Low (0.20–0.40), Moderate (0.40–0.60), High (0.60–0.80), and Very High (0.80–1.00). The complementary 2026 LULC assessment situates this surface within a strongly urbanized landscape: built-up land dominates at 454.8 km² (58.7%), followed by vegetation at 222.3 km² (28.7%), water bodies at 54.6 km² (7.1%), and bare land at 39.8 km² (5.1%), for a total mapped area of 771.5 km² figures consistent with Lahore's well-documented long-term urbanization trajectory (Tahir et al., 2025).

This built-up dominance highlights the need to embed flood susceptibility considerations within land-use and infrastructure planning,

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

complementing earlier AHP-based vulnerability assessments (Zia et al., 2021) and rainwater-harvesting suitability analyses (Mehmood et al., 2025) for the district. The LightGBM susceptibility framework offers a practical decision-support tool for prioritizing drainage, land-use regulation, and green infrastructure, while the LULC assessment underscores the importance of managing urban expansion and protecting vegetation that supports stormwater regulation. Continued LULC updating, improved flood inventories, independent validation via SHAP (Lundberg & Lee, 2017), and integration of exposure and vulnerability indicators (Khan et al., 2026) can further strengthen future flood-risk assessment for Lahore and support evidence-based, climate-resilient urban development. The spatial configuration of the LightGBM susceptibility zones demonstrates an exceptional alignment with the empirical 2025/2026 flood inundation footprints. High and Very High susceptibility classes (index range: 0.60–1.00) are heavily concentrated along the low-lying alluvial plains of Ravi Town and the highly impervious, densely built-up cores of Data Gunj Bakhsh and Gulberg Towns. Consistent with standard machine learning practices, the predictive performance of the trained LightGBM model was rigorously evaluated using the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), overall classification accuracy, and the Cohen's Kappa coefficient. Feature interpretability was executed via relative variable contribution metric scores to dissect the structural weights of the spatial conditioning factors. Based on the validation dataset, the proposed leaf-wise gradient boosting framework achieved an exceptional **AUC-ROC score of 0.994**, indicating near-perfect discriminatory capabilities in mapping local flood susceptibility across the urban landscape. This tight spatial correlation confirms that the proposed leaf-wise gradient boosting approach successfully captures the non-linear interaction between anthropogenic land cover modification and local topographic depressions.

Data Availability

The study utilizes spatial datasets and derived geospatial layers for flood susceptibility and LULC assessment. The final modelling database consists of spatially referenced flood/non-flood samples and extracted conditioning-factor values, and the 2026 LULC assessment is summarized using the four land-cover classes reported above. Underlying raster and vector datasets are available from the corresponding author on reasonable request.

Conflict of Interest

The authors declare no conflict of interest associated with this study.

Acknowledgements

The authors acknowledge the remote sensing, GIS, and geospatial data resources used in developing this analytical framework, and the broader published literature on flood susceptibility and land-use change in Lahore District Metropolitan Area and Pakistan that informed the interpretation of results.

REFERENCES

Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). *A Land Use and Land Cover Classification System for Use with Remote Sensor Data* (U.S. Geological Survey Professional Paper 964). U.S. Government Printing Office.

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin, 24*(1), 43–69.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

Bui, D. T., Pradhan, B., Nampak, H., Bui, Q.-T., Tran, Q.-A., & Nguyen, Q.-P. (2016). Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modeling in a high-frequency tropical cyclone area. *Journal of Hydrology, 540*, 317–330.

Chapi, K., Singh, V. P., Shirzadi, A., Shahabi, H., Bui, D. T., Pham, B. T., & Khosravi, K. (2017). A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environmental Modelling & Software, 95*, 229–245.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment, 37*(1), 35–46.

Costache, R., Pham, Q. B., Sharifi, E., Linh, N. T. T., Abba, S. I., Vojtek, M., Vojteková, J., Nhi, P. T. T., & Khoi, D. N. (2020). Flash-flood susceptibility assessment using multi-criteria decision making and machine learning supported by remote sensing and GIS techniques. *Remote Sensing, 12*(1), 106.

Ibrahim, M., Huo, A., Ullah, W., Ullah, S., & Xuantao, Z. (2025). An integrated approach to flood risk assessment using multi-criteria decision analysis and geographic information system: A case study from a flood-prone region of Pakistan. *Frontiers in Environmental Science, 12*, 1476761. https://doi.org/10.3389/fenvs.2024.1476761

Islam, A. R. M. T., Talukdar, S., Mahato, S., Kundu, S., Eibek, K. U., Pham, Q. B., Kuriqi, A., & Linh, N. T. T. (2021). Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers, 12*(3), 101075.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30*, 3146–3154.

Khan, M., Chen, R., & Khan, S. (2026). Integrated flood susceptibility and multi-temporal flood risk prioritization in Pakistan using hydro-climatic and geospatial indicators. *Hydrology, 13*(7), 170. https://doi.org/10.3390/hydrology13070170

Latif, R. M. A., & He, J. (2025). Flood susceptibility mapping in Punjab, Pakistan: A hybrid approach integrating remote sensing and analytical hierarchy process. *Remote Sensing*.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*.

Mahmood, A., Nasreen, S., Nawaz, M. A., Khalid, A., & Ali, A. (2019). Assessment of urban floods in Lahore city using remote sensing and GIS techniques. *Journal of the Indian Society of Remote Sensing, 47*(7), 1199–1209. https://doi.org/10.1007/s12524-019-01032-1

Mahmood, S., & Hamayon, K. (2021). Geo-spatial assessment of community vulnerability to flood along the Ravi River, Ravi Town, Lahore, Pakistan. *Natural Hazards*.

ISSN (Online): 3006-2071 | ISSN (Print): 3006-2063
Volume 4 • Issue 1 • March 2026

Majeed, M., Lu, L., Anwar, M. M., Tariq, A., Qin, S., El-Hefnawy, M. E., El-Sharnouby, M., Li, Q., & Alasmari, A. (2023). Prediction of flash flood susceptibility using integrating analytic hierarchy process (AHP) and frequency ratio (FR) algorithms. *Frontiers in Environmental Science*.

Mehmood, A., Basheer, M. A., Arshad, H. S. H., Zia, S., Muntaha, S. T., & Riaz, A. R. (2025). Assessing the potential of rainwater harvesting through GIS and remote sensing techniques in combating urban flooding in Lahore, Pakistan. *GeoJournal, 90*, Article 197.

Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R., & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews, 207*, 103225.

Mosavi, A., Ozturk, P., & Chau, K.-W. (2018). Flood prediction using machine learning models: Literature review. *Water, 10*(11), 1536.

Pham, B. T., Luu, C., Van Phong, T., Nguyen, H. D., Le, H. V., Tran, T. Q., Ta, H. T., & Prakash, I. (2020). Flood risk assessment using deep learning integrated with multi-criteria decision analysis. *Knowledge-Based Systems, 219*, 106899.

Rahimi, M., Malekmohammadi, B., Karimi Firozjaei, M., Kerachian, R., Jokar Arsanjani, J., Tan, M. L., Awange, J., Savic, D., Duan, Q., & AghaKouchak, A. (2026). Integrating geospatial intelligence and machine learning for flood susceptibility mapping. *Scientific Reports, 16*, Article 10228. https://doi.org/10.1038/s41598-026-41014-3

Rentschler, J., Salhab, M., & Jafino, B. A. (2023). Global evidence of rapid urban growth in flood zones since 1985. *Nature, 622*(7981), 87–92. https://doi.org/10.1038/s41586-023-06468-9

Saikh, A., & Mondal, P. (2023). Flood susceptibility mapping using GIS and machine learning: A case study of Eastern India. *Environmental Science and Pollution Research*.

Tahir, Z., Haseeb, M., Mahmood, S. A., Batool, S., Abdullah-Al-Wadud, M., Ullah, S., & Tariq, A. (2025). Predicting land use and land cover changes for sustainable land management using CA-Markov modelling and GIS techniques. *Scientific Reports, 15*, Article 3271. https://doi.org/10.1038/s41598-025-87796-w

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology, 512*, 332–343.

Waleed, M., & Sajjad, M. (2025). High-resolution flood susceptibility mapping and exposure assessment in Pakistan: An integrated artificial intelligence, machine learning and geospatial framework. *International Journal of Disaster Risk Reduction, 121*, Article 105442.

Wang, Y., Fang, Z., Hong, H., & Peng, L. (2020). Flood susceptibility mapping using convolutional neural network frameworks. *Journal of Hydrology, 582*, 124482.

Zia, S., Shirazi, S. A., & Nasar-u-Minallah, M. (2021). Vulnerability assessment of urban floods in Lahore, Pakistan using GIS based integrated analytical hierarchy approach. *Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences, 58*(1). https://doi.org/10.53560/PPASA(58-1)604