

Interpretable Bi-LSTM Based Model for Understanding Urban Growth Patterns Using Explainable AI

Jambukeshwar Pujari

pujarikeshwar@gmail.com

KIT's College of Engineering

Vanishree Kavi Mahesh

Mangalayatan University

Research Article

Keywords: Urban Growth Modeling, Bi-LSTM, Explainable AI (XAI), Urban Sustainability Score, SMOTE, Feature Selection, Deep Learning, Land Use, Urban Planning

Posted Date: April 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9072147/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Interpretable Bi-LSTM Based Model for Understanding Urban Growth Patterns Using Explainable AI

^{1*}Jambukeshwar Pujari, ²Vanishree Kavi Mahesh

Abstract Urban growth has emerged as one of the most daunting planning issues in rapidly expanding metropolitan regions, spurred by an intricate interdependence of socio-economic, environmental, and spatial considerations. In this investigation, it is suggested that an interpretable deep learning model that combines Bidirectional Long Short-Term Memory (Bi-LSTM) networks with Explainable AI (XAI) methods be used to model and analyze urban growth patterns. Taking advantage of a high-dimensional Sustainable Urban Planning and Landscape dataset, the model is preprocessed with state-of-the-art preprocessing involving class balancing using SMOTE, ensemble feature selection, and dimensionality reduction by truncated SVD. The Bi-LSTM model is fine-tuned by the Hybrid Hiking with Dollmaker Optimization Algorithm (HDOA) to forecast Urban Sustainability Scores (0–1). SHAP is employed for transparent visual explanation of feature contributions and decision paths. Along with better accuracy in predictions, the proposed method identifies prominent drivers of urban growth in green cover, population density, and connectivity through roads. Comparative analysis with recent hybrid models of literature brings out the robustness, explainability, and policy-relevance of our model. The insights drawn from this study provide scalable and actionable intelligence for urban planners and sustainability experts.

Keywords: Urban Growth Modeling; Bi-LSTM; Explainable AI (XAI); Urban Sustainability Score; SMOTE; Feature Selection; Deep Learning; Land Use; Urban Planning

^{1*}Prof. Jambukeshwar Pujari,
Assistant Professor,
Department of Computer Science and Business Systems,
KIT's College of Engineering,
Kolhapur, Maharashtra, 416234 India
Email: pujarikeshwar@gmail.com

²Mrs. Vanishree Kavi Mahesh
Research Scholar
Department of Computer Science and Engineering
Mangalayatan University
Aligarh, Uttar Pradesh, 202146 India
Email: vanishreekavi@outlook.com

*Corresponding Author: Jambukeshwar Pujari
Email: pujarikeshwar@gmail.com

1. Introduction

Urbanization is the growth of cities fueled by intensified human activities and influenced by economic, social, political, and geographical factors (Bharath et al. 2018). Historically considered to be one of the major drivers of economic development, the focus in recent decades has been on attaining sustainable urbanization that is still one of the most important challenges for developing nations. Millions of individuals in most of these countries remain without access to basic services like electricity, clean water, and proper sanitation facilities (Castells-Quintana 2017). The intersection of a fast rate of urban growth and increasing urban populations puts immense pressure on scarce land resources, commonly leading to comprehensive environmental degradation. Urbanization introduces tremendous land-use and land-cover changes—mainly as a result of population growth and related anthropogenic pressures—accruing further to global climate change (Wang and Wang 2017). To better control urban regions, it is important to comprehend the urban growth dynamics and its attendant environmental implications. In response to this demand, urban expansion modeling has become a central scientific tool to study and forecast the temporal and spatial development of cities (Sathish Kumar et al. 2013). The impact of human activities on environmental change has been a core interest of land management scientists for many years. Such changes, which are mostly spatial in scope, are assessed using qualitative and quantitative methods. Yet, recording the spatial and temporal heterogeneity of urban and rural landscapes is still a sophisticated task, mainly because of traditional historical constraints in spatial analysis technologies, which have not yet been adequate enough to meet the actual real-time planning and decision-making requirements of urban managers and policymakers. Old-fashioned land-use planning techniques—e.g., field surveys and participatory mapping are labor-demanding, time-consuming, and expensive. Thankfully, improvements in data acquisition technologies and the rise in computing power have enabled us to implement algorithms previously thought to be purely theoretical (Samardžić-Petrović et al. 2017). Remote Sensing (RS) and Geographic Information Systems (GIS) technologies form the core of land use visualization, monitoring, and interpretation. RS data are specifically key to the monitoring of land use and land cover changes and provide coverage ranging from regional to global levels. Analyzing these datasets provides insights into both natural and built environments. Over the past two decades, significant technological improvements in RS sensors and techniques have led to the availability of high-quality data with enhanced spatial resolution (Cao et al. 2019). The fusion of high-resolution data from sources like LIDAR, RADAR, MSS, hyperspectral imagery, UAVs, and other airborne and satellite platforms has considerably enhanced the accuracy and intensity of land use estimations. In contrast to traditional methods where spatial and non-spatial data were usually kept apart, the present fusion of RS and GIS tools facilitates the incorporation of socio-economic variables enabling a comprehensive understanding of urban processes and more informed land-use planning.

Land use planning per se is an orderly process meant to achieve the optimal and sustainable utilization of land for the well-being of people. In order to effectively manage land resources, planners and policymakers must put in place regulation systems such as land use regulations, zoning legislation, allocation strategies, and control policies that enhance sustainability and resource

preservation. But with the fast pace of urbanization, the development tends to be spontaneous and random, and hence there are problems of land degradation, rising air pollution, and water resource contamination. Among the biggest challenges for planners and policymakers is guiding this development in an organized and sustainable way. Planned development relies very much on the availability of spatial, non-spatial, and temporal data. When merged with socio-economic data, EO and GIS data are strong means for redrawing and enhancing land use maps, which are necessary for addressing the challenges of uncontrolled urban expansion. In remote sensing, two main parameters—spatial resolution and temporal frequency—are important for examining land use change. Both proprietary and open-source EO and GIS data are supplied by different national and global organizations. Land use planners need detailed spatial information, such as built-up and non-built-up area classifications, vegetation cover, urban forests, water bodies, drainage systems, and transportation networks. Much of the data is available directly from remote sensing imagery. Open-access data platforms like OpenStreetMap, USGS Earth Explorer, Sentinel Hub, Copernicus Open Access Hub, and the Global Human Settlement Layer (GHSL) provide useful repositories of spatial information. In addition, commercial satellite imagery, aerial images, and LIDAR data find extensive usage in high-end applications like cadastral mapping, 3D modeling, and advanced feature extraction. Land use planning is usually carried out in a process of three steps: initially, through the examination of past data—like census reports, master plans, and statistical data from different government and non-governmental organizations; secondly, by determining to what level existing development meets previous planning estimates, primarily by utilizing EO data for cross-validation; and finally, by predicting future land use requirements based on anticipated statistical trends and government policy systems (Chaturvedi and de Vries 2021).

Urbanization, as a worldwide phenomenon, is still transforming the form and operation of cities globally. Urban cities are not just crucial to national economies but are becoming increasingly the major drivers of economic development. In the Indian scenario, about 31% of people live in cities, and they contribute about 63% to the national GDP (Census 2011). According to projections, by the year 2030, almost 40% of India's population will reside in urban areas, contributing to about 75% of India's GDP (Khan et al. 2020). Housing this swelling population demands integration planning that encompasses physical, social, economic, and institutional infrastructure. An indispensable step towards this is the potential to foresee and recognize the new areas of urban expansion. Cities are being transformed globally through both natural and anthropogenic impulses. In the last two and a half centuries, cities have been the major target of anthropogenic forces. Urbanization, which gained momentum with industrialization, largely changed old systems of production and markets. This change promoted population increase and concentration in urban centers—places that support agricultural, industrial, distribution, governmental, cultural, and social activities. With the development of cities, they turned into multifaceted nodes where different economic activities and social agents converse within different spatial scales (Cengiz et al. 2019). The industrial period also transformed cities as the focal points of intense economic, cultural, and social interaction. At the same time, the mechanization of agriculture minimized rural labor requirements, leading to extensive rural-urban migration. The process has significantly impacted urban development patterns

and land-use planning decision-making by professionals and stakeholders (Aithal and Ramachandra, 2020). Consequently, cities have been experiencing ongoing and fast-paced spatial change, resulting in a variety of challenges including environmental degradation, socio-economic disparities, and governance concerns. These encompass unequal access to urban infrastructure and services, together with the encroachment of agricultural and natural land on the urban fringes into developed land (Hou et al. 2020; Zhou et al. 2019).

By the 1980s, urban spatial growth had become an issue of concern, with scholars differentiating between two overarching forms: urban sprawl and compact development. Though both have been studied for their respective advantages and disadvantages, studies highlight that urban sprawl invariably leads to more adverse environmental and social effects compared to compact development. Urban expansion, sprawling or compact, is determined by economic and demographic forces above all. However, its effects spill over into other socio-spatial and political-economic arenas. These involve crowding, increased land value, slums, high crime rates, and unequal distribution of public services. Additionally, the environmental impacts are considerable—urban growth changes soil formations, damages biodiversity, impacts energy flows, and shapes local and regional climate regimes. Hence, it is necessary to conduct a detailed analysis of its extent, direction, and temporal dynamics. No less significant is also examining those cultural, ecological, and political-economic processes that influence and are influenced by urban growth. It is crucial to define these dynamics clearly for planners, decision-makers, and resource managers in charge of adopting sustainable urban management strategies (Cengiz et al. 2022). Traditional urban growth modeling methods, including cellular automata (CA), agent-based models (ABMs), and statistical regression methods, have played an important role in describing spatial dynamics. These methods tend to lack capturing temporal dependency and intricate non-linear relationships in the data.

Recent breakthroughs in deep learning, especially in Recurrent Neural Networks (RNNs), along with its variants, such as Bidirectional Long Short-Term Memory (Bi-LSTM) networks, have indicated potential in processing temporal and sequential data, making them beneficial for the modeling of spatio-temporal patterns of urban development. While powerful predictors, deep learning models are often slammed for their "black-box" nature, which denies explainability and discourages their use in policy-sensitive decision-making. The increased demand for transparency and accountability for AI-based systems gave rise to Explainable Artificial Intelligence (XAI) approaches. XAI methodologies seek to offer insights understandable to humans regarding model behavior and decisional logic, and with them, it is feasible to track which features or patterns contribute to urban expansion forecasts. Here, the present research introduce an interpretable Bi-LSTM based model combined with Explainable AI methods to reason and learn urban expansion patterns. The use of the Bi-LSTM model captures forward and backward temporal relationships of urban data to enhance learning from urban history. In order to increase transparency, we utilize XAI tools like SHAP (SHapley Additive exPlanations) and attention mechanisms to explain the model's predictions and emphasize the most influential driving features of urban growth. Through the integration of temporal deep learning and explainability, the suggested model not only forecasts

upcoming urban growth with high accuracy but also offers actionable explanations of the causes of urban transformation. The major contributions of the current research are:

- Suggested a deep stacked Bi-LSTM approach to learn intricate spatial, temporal, and socio-economic interactions for Urban Sustainability Score prediction.
- Implemented SHAP to offer interpretable, global, and local model explanations—allowing urban planners to know why predictions were made.
- Utilized RFE, LASSO, and Boruta algorithms and Truncated SVD to decrease feature dimensionality and improve model stability.
- Applied Hybrid Hiking with Dollmaker Optimization Algorithm (HDOA) to optimize Bi-LSTM parameters for enhanced learning effectiveness and prediction accuracy.
- Created a scalable model design that offers actionable insights, facilitating sustainable urban planning on a range of city datasets.

The structure of this paper is as follows: Section 2 involves a comprehensive literature review and the definition of current gaps in the modeling of urban growth. Section 3 introduces the data, which consists of major spatial, environmental, and socio-economic indicators related to urban sustainability. Section 4 presents the introduced interpretable Bi-LSTM model, where preprocessing, feature selection, model architecture, and the explainability layer are explained. Section 5 outlines the experimental configuration, performance assessment, and SHAP explainability outcomes. Section 6 reflects on the implications of the findings in the sustainable urban planning context. Lastly, Section 7 summarizes the paper and presents future work directions.

2. Literature review

Vaibhav Kumar et al. (2021) proposed a hybrid approach that integrated Agent-Based Modeling (ABM) and Cellular Automata Markov (CAM) to model land-use and land-cover (LULC) transformations for urban expansion. The model, when applied in Dehradun, India, attained 87.86% spatial accuracy, depicting massive conversion of agricultural land and vacant land to residential and commercial use. CAM identified spatial-temporal shifts, and ABM indicated human and economic drivers. The combined model performed better than isolated CAM both in prediction and intuition, albeit its scalability being restricted by the availability of socio-economic and behavioral information. Expanding on this union of spatial processes with socio-economic factors, Badri Gari and Rohith (2022) formulated a decision-support system for Hyderabad's Zone-1 LULC planning incorporating hexagonal spatial planning and a vector-CA approach. They used rule-based models and logistic regression to test the impact of urban policies and develop suitability maps. The maps, together with key drivers, were used in an agent-based model to develop high-growth areas and best locations for housing parcels. The method, being effective in spatial analysis, may not be able to respond to changes in real-time socio-political dynamics or abrupt changes in regulatory policies.

Building on this idea of integrating spatial simulation with behavioral modeling, Najmeh and Tõnu (2021) came up with a CA-Agent model by merging CA, ABM, and the Markov model to simulate intra-urban complexities and spatial transitions. In their work, every cell acts as an active agent, with results defined by behavioral rules as well as neighboring impacts. Validated against

2018 data (Kappa = 0.86), the model forecasted urban development to be 175.24 sq. km by 2030, a 30.25% increase from 1990. Although good at modelling land-use change and emphasizing the value in keeping agricultural and forested land intact, its fixed probability of transition restricts responsiveness to unforeseen socio-economic change. Solving the problem of needing a more flexible and data-fidelity prediction, Gao et al. (2025) proposed the Urban Expansion Scenario Prediction (UESP) model with an agent-based simulation and multi-head attention mechanism to predict urban development more accurately. It integrates both macro-pattern and micro-human activities and achieves a high prediction accuracy of 0.925, surpassing traditional models. When applied to the Beijing–Tianjin–Hebei region, it predicted the expansion of construction land mainly in core cities, while forest and farmland extended in ecological protection plans. The performance of the model relies on high-density, high-quality POI and behavioral information, limiting its applicability in low-data availability areas.

Going further into integrating deep learning for spatial-temporal prediction, Xinhao Pan et al. (2022) created a sophisticated Urban Expansion (UE) model that integrated U-Net and LSTM to forecast urban expansion in the Beijing–Tianjin–Hebei region. The model was highly performed with 99.18% accuracy, Kappa of 0.88, and FoM of 0.13 by relying on U-Net's spatial feature extraction and LSTM's temporal modeling. It estimated urban land growth to 8736–9155 km² in 2039–2043, or 10.99–16.31% higher than in 2020. Although full of promise for urban planning and other uses, the success of the model is contingent upon the existence of high-quality time-series spatial data. To deal with the issue of heterogeneous spatial data formats, Qingfeng Guan et al. (2024) presented the HashGAT-Vector CA (VCA) model for simulating urban land use evolution. It converts irregular parcel data into fixed-length vectors with a hash function and represents spatial relationships in a graph structure. A Graph Attention Network (GAT) extracts environmental effects and parcel interactions to forecast land use transformation. Implemented in Shenzhen (2009–2012), the model dominated other VCA models and mapped scenarios for 2025 and 2030 under ecological scenarios. Though its usability relies on correct spatial topology, thereby restricting applications to places with poorly delineated or dynamic parcel boundaries.

Following the trend of CA model improvement toward more realistic urban sprawl simulation, Xiaoyong Tan et al. (2023) proposed the Spatial Hierarchical Learning-based CA (SH-CA) model to resolve spatial-temporal mismatches. Redefining neighborhood patterns and employing lightweight CNNs with a gated filter mechanism, the model is successful in capturing intricate spatial-temporal interactions. When applied to Beijing, Guangzhou, and Chengdu (2010–2015), SH-CA was superior to four other CA models in terms of realism and figure of merit. Hierarchical learning module was found to be flexible in different CA systems, albeit with the cost of complexity and careful tuning with respect to well-defined multiscale neighborhoods and convolutional parameters for varying urban environments. Compared to the above, emphasizing long-term urban sprawl analysis with more traditional approaches, Najib Ansari et al. (2024) investigated land use/land cover changes in Kolkata Urban Agglomeration from 1991 to 2021 and forecasted trends to 2100 with Landsat data and a CA-Markov model in TerrSet. The research indicated high urban sprawl with decreasing vegetation and agricultural fields, with central Kolkata undergoing population loss due to out-

migration. With kappa accuracy of 0.8, the model enables sustainable urban planning. Nevertheless, dependence on unsupervised classification (k-means) tends to be less accurate than supervised methods.

Returning to sophisticated deep learning architectures, Sam Navin et al. (2022) proposed a Vision Transformer–Bi-LSTM model for LULC transformation prediction in Javadi Hills, India, using LISS-III and Landsat data. The model achieved high accuracy of classification (98.76%) and strong validation (98.38%) through the support of land surface temperature and LULC maps. Google Earth Engine was used to interpret the results and help planners as well as forest authorities. Its performance relies heavily on the quality of satellite data and cloud infrastructure. In view of the significance of not just accuracy but interpretability too, Zelin Wang et al. (2025) integrated eXplainable AI (XAI) into ML/DL models for more interpretable land-use change projections. Their approach was extremely accurate and computationally efficient with explanations of key drivers of land-use change, such as vacant lots and neighborhood dynamics. The method assists planners and researchers in understanding drivers of city growth better but does not work for exceptional or unexpected land-use transformation not occurring in the past.

Additional increasing explainability of predictive modeling, Minjun and Geunhan (2022) used an XAI-based approach by combining XGBoost and SHAP values to predict urban expansion in South Korea using land cover data for 2007 and 2019 and a set of influencing it. They found that proximity to developed regions and lowland increase growth, whereas mountains and land-use regulations are hindrances to it. The model positively identifies the most powerful drivers of urbanization but maybe cannot generalize well for places with different geographic or policy contexts. Taking this combination of interpretability and fine-scale forecasting one step ahead, Yue Zhu et al. (2024) presented an ensemble approach with DL-based ensemble framework for the simulation of fine-scale urban growth leveraging transformers for multi-temporal and CNNs for single-temporal spatial feature extraction. A channel-wise attention mechanism was added in order to improve interpretability by highlighting the significance of features. The model was more spatially and temporally accurate than competing approaches in the test on Shenzhen but could be hampered by high complexity and data requirements for scalability.

Summing up these improvements, Baoling Gui et al. (2025) developed a U-Net++-based approach that combined attention mechanisms and autoregressive CA to simulate urban expansion with Changsha's spatiotemporal dataset (2010–2022). With accuracy of 0.87 and a Figure of Merit of 0.90, the model outperformed traditional techniques in representing spatial and temporal variation remarkably well. Its end-to-end structure ensures adaptive and uniform urban development predictions, with potential for global applicability. However, in line with previous findings from several studies, the shortcoming of the current model is that it fails to combine dynamic policy and socio-political inputs, suggestive of a common deficiency among both rule-based and AI-based LULC modeling techniques. The existing literatures summary is shown in table.1.

Table. 1: Summary of existing literatures

Ref.	Authors	Model /	Case Study	Key	Limitations
------	---------	---------	------------	-----	-------------

		Approach Used	Area	Contributions	
[14]	Vaibhav Kumar et al.	Hybrid CAM + ABM	Dehradun, India	Integrates socio-economic factors and stakeholder behavior for LULC transition modeling	Depends on rich socio-economic and behavioral datasets; limited regional scalability
[15]	Badri Gari & Rohith	Vector-Cellular Automata + Logistic Regression + ABM	Hyderabad, India (Zone-1)	Hexagonal spatial planning with decision support for housing development	Cannot capture sudden socio-political or policy changes
[16]	Najmeh & Tõnu	CA-Agent Model (CA + ABM + Markov)	Not specified	Simulates spatial heterogeneity and agent interactions; 30% urban growth forecasted	Relies on preset transition probabilities; reduced flexibility to abrupt changes
[17]	Gao et al.	Urban Expansion Scenario Prediction (UESP)	Beijing–Tianjin–Hebei, China	Uses multi-head attention + ABM; high accuracy (0.925); models macro and micro urban dynamics	Requires high-density POI and behavioral data; limited in data-scarce environments
[18]	Xinhao Pan et al.	U-Net + LSTM Neural Network	BTHUA, China	99.18% accuracy; strong spatio-temporal modeling with deep learning	Needs high-quality time-series data; limits widespread deployment
[19]	Qingfeng Guan et al.	HashGAT + Vector Cellular Automata	Shenzhen, China	Graph Attention Network + vectorization of land parcels; improves urban LULC simulation accuracy	Dependent on clear and accurate parcel boundary/topology data

[20]	Xiaoyong Tan et al.	Spatial Hierarchical Learning Cellular Automata (SH-CA)	Beijing, Guangzhou, Chengdu	Models multiscale neighborhood interactions; better realism than traditional CA models	Requires fine-tuned convolution parameters and well-defined neighborhood structures
[21]	Najib Ansari et al.	CA-Markov + K-means Clustering	Kolkata Urban Agglomeration	Projects LU/LC to 2100; strong accuracy (kappa = 0.8); highlights sprawl and vegetation decline	Uses unsupervised classification (k-means), which may lower mapping accuracy
[22]	Sam Navin et al.	Vision Transformer + Bi-LSTM	Javadi Hills, India	High-accuracy LULC change prediction using LISS-III and Landsat data; validation using LST and classification maps; GEE integration for interpretation	Depends on high-quality multispectral data and cloud-based resources
[23]	Zelin Wang et al.	ML/DL models integrated with Explainable AI (XAI)	Not specified	High prediction accuracy; enhanced interpretability of land-use drivers via XAI; useful for policymakers	Struggles to capture rare or abrupt land-use changes not present in historical data
[24]	Minjun & Geunhan	XGBoost with SHAP values (XAI-based)	South Korea	Identified key urban growth drivers; highlighted terrain and regulatory impacts; ranked	Limited generalizability across different geographic or policy contexts

				factor importance	
[25]	Yue Zhu et al.	Deep learning ensemble: Transformers + CNNs + Channel-wise Attention	Shenzhen, China	Fine-scale urban simulation; high spatial/temporal accuracy; improved interpretability through attention modules	Requires high computational power and large training datasets
[26]	Baoling Gui et al.	U-Net++ with Attention Mechanisms + Autoregressive Cellular Automata	Changsha, China	Achieved strong accuracy (OA 0.87, FoM 0.90); robust handling of spatiotemporal heterogeneity; scalable end-to-end design for prediction	Lacks integration of dynamic policy and socio-political variables

3. Research gap

Even with prominent developments in urban growth and LULC modeling, there are salient areas of research deficiency. Some models like those by Vaibhav Kumar et al. (2021) and Badri Gari & Rohith (2020) emphasize the incorporation of socio-economic factors but are hampered by reliance on region-specific information and rule-based static assumptions, preventing flexibility to dynamic urban contexts. Most models, such as the CA–Agent model (Mozaffaree Pour and Oja, 2021) and SH-CA model (20), depend heavily on pre-defined transition probabilities or convolutional parameters, which need to be accurately tuned and limit model adaptability when faced with abrupt socio-political or economic shifts. Deep learning-based models like Xinhao Pan et al. [2022] and Qingfeng Guan et al. (2024) provide high prediction accuracy at the cost of interpretability, rendering them less usable for policymakers and planners. High data demands like dense POI data, behavior patterns, or well-mapped spatial topologies also limit model scalability, as observed in the UESP (Gao et al. 2025) and HashGAT-VCA [19] models. In addition, unsupervised classification techniques (e.g., k-means employed by Najib Ansari et al. (2024) can be at the expense of baseline LU/LC accuracy compared to more resilient supervised techniques. These constraints highlight an obvious demand for data-efficient, interpretable, and generalizable models that can naturally

integrate both macro-level land transformation and micro-level human activity while being resilient to data variability and policy change.

4. Proposed Methodology

This study suggests a strong, explainable, and scalable deep learning architecture for analyzing urban morphology, aiming to predict Urban Sustainability Scores from diverse spatial, socio-economic, and environmental data. The approach couple's thorough data preprocessing, sophisticated feature selection, with a deep stacked Bidirectional Long Short-Term Memory (Bi-LSTM) network optimized using the Hybrid Hiking with Dollmaker Optimization Algorithm (HDOA). The architecture is also enriched with explainable AI (XAI) tools most notably SHAP for interpretability and real-world decision-making. Overall pipeline, shown in Fig. 1, starts with data collection from the Sustainable Urban Planning & Landscape Dataset and continues with stringent cleansing, normalization, encoding, and imbalance treatment. The dimensionality is reduced with Truncated SVD, and an ensemble feature selection method (RFE, LASSO, Boruta) is used to get the most knowledge-rich attributes. The cleaned dataset is further employed in training a deep Bi-LSTM network that learns nonlinear and temporal relationships between urban features. The hyperparameters of the model are optimized using HDOA for max learning efficiency and generalization. Lastly, the predictions are explained by interpreting the output via SHAP values to make it transparent, trustworthy, and domain-relevant for urban planning stakeholders, shown in Fig. 1 given below.

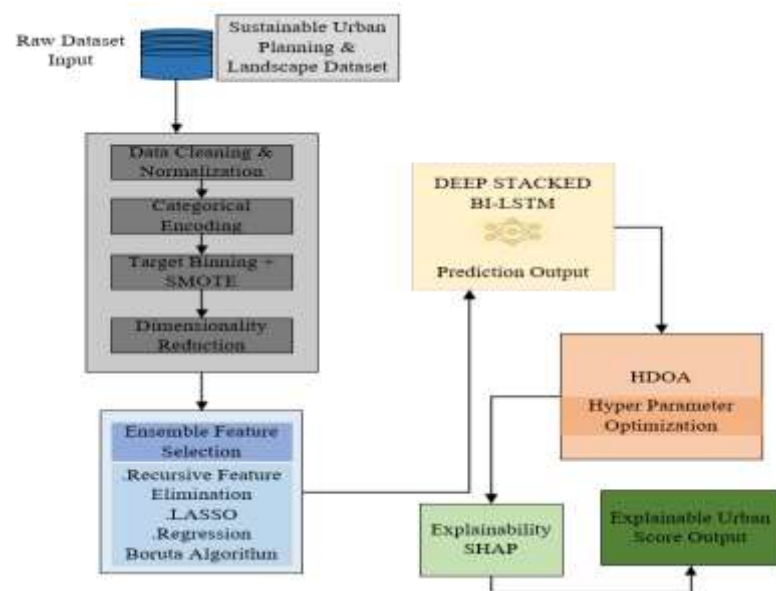


Fig. 1: Methodology Flow

4.1 Data Collection

The data used in this research, Sustainable Urban Planning & Landscape Dataset (Sustainable Urban Planning & Landscape Dataset), is obtained from Kaggle and openly shared under a CC0 (Public Domain) license. The dataset comprises 3,476 rows of city-level observation with 16 urban,

environmental, and socio-economic variables essential for urban sustainability modeling. Data are assumed to be drawn from actual urban smart city infrastructure and city databases, though no particular source institutions or time stamps are explicitly mentioned. The variable of interest is a normalized Urban Sustainability Score (0 to 1), which is a composite representation of a city's ecological, spatial, and social sustainability. The data contain class imbalance in score distribution, which needs preprocessing measures such as SMOTE during modeling. The major features of the dataset are summarized in Table 2, which is provided below.

Table 2: Key Features in the Dataset

Feature Category	Feature Name	Explanation
Urban Spatial Data	building_density	Ratio of built-up area to total land
	road_connectivity	Measure of urban road network connectivity
	public_transport_access	Index of public transport proximity and availability
Environmental Data	air_quality_index	Composite score for air pollution levels
	green_cover_percentage	Percentage of green land in the area
	carbon_footprint	Estimated CO ₂ emissions per capita
Socio-Economic Data	population_density	Number of people per square km
	crime_rate	No. of stated crimes per 1,000 residents
	avg_income	Average income level in the city
Sustainability & Risk	renewable_energy_usage	Share of energy consumption from renewable sources
	disaster_risk_index	Vulnerability to natural disasters
Target Variable	urban_sustainability_score	Composite score ranging from 0 to 1

4.2 Data Preprocessing

The raw dataset encompassing diverse spatial, socio-economic, and environmental attributes requires rigorous preprocessing before being ingested by the Bi-LSTM model. The preprocessing pipeline includes cleansing, normalization, categorical encoding, imbalance handling, and dimensionality reduction. Each step is described below:

4.2.1 Data Cleansing and Normalization

Initial preprocessing involves handling missing values and outliers. Missing entries in numerical features were imputed using feature-wise median imputation, and outliers were detected using z-score normalization, given in Eqn. (1)

$$Z_i = \frac{X_i - \mu}{\sigma} \quad (1)$$

where X_i is the original value, μ is the mean, and σ is standard deviation. Any values outside the threshold of $|Z_i| > 3$ were considered outliers and replaced by the 95th percentile of that feature.

Next, min-max normalization was applied to bring all features into the [0,1] range, Eqn. (2)

$$X_i^{scaled} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

This ensures convergence during training and prevents gradient instability in LSTM models.

4.2.2 Encoding Categorical Variables

The dataset comprises a categorical feature: `land_use_type`. This was transformed using one-hot encoding using Eqn. (3), which creates binary columns for each category $c_k \in \{C_1, C_2, \dots, C_K\}$, where K is the total amount of unique land use types:

$$x^{(k)} = \begin{cases} 1 & \text{if the instance belongs to category } C_k \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

This encoding ensures the model does not impose ordinal relationships between nominal values.

4.2.3 Handling Imbalanced Target Distribution

Since the target variable `urban_sustainability_score` is continuous in nature (ranging from 0 to 1), direct application of SMOTE is not valid. Therefore, the target was discretized into three ordinal bins using equal width binning, using Eqn. (4)

$$Class(y) = \begin{cases} Low & \text{if } 0 \leq y < 0.33 \\ Medium & \text{if } 0.33 \leq y < 0.66 \\ High & \text{if } 0.66 \leq y \leq 1 \end{cases} \quad (4)$$

Then, SMOTE was applied to generate synthetic samples for underrepresented classes using Eqn. (5)

$$x_{new} = x_i + \lambda \cdot (x_j - x_i), \lambda \sim U(0,1) \quad (5)$$

where x_i is a minority sample, x_j is one of its k -nearest neighbors, and λ is a random scalar drawn from a uniform distribution.

4.2.4 Dimensionality Reduction

To mitigate the curse of dimensionality and reduce computation during model training, Truncated SVD [Di Wu, 2024] was applied using Eqn. (6)

$$X \approx U_k \sum_k V_k^T \quad (6)$$

Where, $X \in \mathbb{R}^{n \times d}$ is the high-dimensional feature matrix, $U_k \in \mathbb{R}^{n \times k}$ contains the left singular vectors, $\sum_k \in \mathbb{R}^{k \times k}$ is a diagonal matrix of the top- k singular values, $V_k \in \mathbb{R}^{k \times d}$ contains the right singular vectors.

The reduced representation $U_k \sum_k$ preserves most of the variance (typically 90-95%) while significantly reducing dimensionality, allowing for efficient LSTM training.

This preprocessing pipeline enables the interpretable Bi-LSTM model to learn from a clean, balanced, and compact dataset, maximizing predictive performance while maintaining generalizability.

4.3 Feature Selection (Ensemble Approach)

The significant dimension and multicollinearity of features in urban datasets can impair deep learning models' performance and interpretability. This was addressed by combining the LASSO

Regression, and RFE (Urmi et al., 2024) and Boruta Algorithm (Manikandan et al., 2024) in an ensemble feature selection method. In order to discover the most useful and non-redundant features, our multi-perspective technique guarantees robustness. Fig. 2 illustrates the ensemble feature selection pipeline combining three distinct methods: RFE (wrapper), LASSO (embedded), and Boruta (filter).

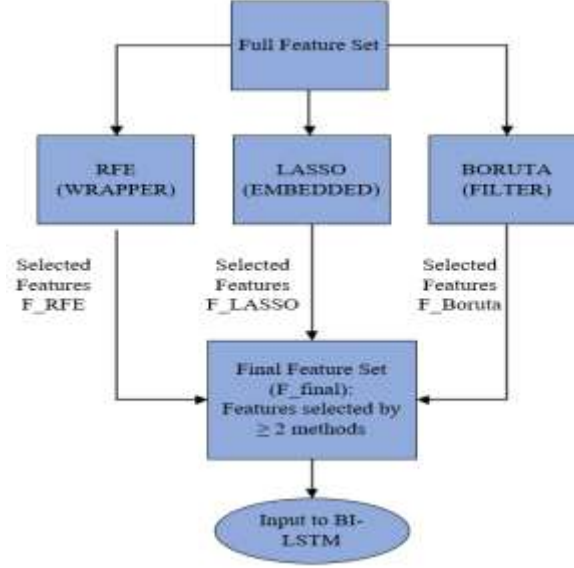


Fig. 2: Ensemble Feature Selection Pipeline

4.3.1 Recursive Feature Elimination (RFE)

Based on the weights of a model, RFE is a wrapper-based technique that iteratively eliminates the least significant features. During every iteration, it:

1. Trains a base model.
2. Computes a feature ranking score.
3. Removes the least important feature.
4. Repeats until the desired number of features k is selected.

Let the trained model output feature importance weights $w = [w_1, w_2, \dots, w_d]$. The feature with the smallest absolute weight is eliminated at each step given in Eqn. (7)

$$\min_{j \in \{1, \dots, d\}} |w_j| \Rightarrow \text{Remove feature } x_j \quad (7)$$

This results in a reduced subset $\{x_{j_1}, x_{j_2}, \dots, x_{j_k}\}$ optimized for the chosen model.

4.3.2 LASSO Regression

LASSO is a regularization-based method that imposes an L_1 penalty on the magnitude of coefficients in a linear regression model. The objective function is given in Eqn. (8)

$$\min_{\beta} \left(\frac{1}{2n} \sum_{i=1}^n (y_i - x_i^T \beta)^2 + \lambda \sum_{j=1}^d |\beta_j| \right) \quad (8)$$

Here, x_i is the feature vector for instance i , y_i is the target, β are the regression coefficients, and the regularization parameter λ .

LASSO forces some coefficients β_j to accurately zero, effectively performing feature selection given in Eqn. (9)

$$\beta_j = 0 \Rightarrow \text{Feature } x_j \text{ is irrelevant and removed} \quad (9)$$

4.3.3 Boruta Algorithm

Boruta is an all-relevant feature selection method built around the random forest algorithm. Unlike LASSO and RFE, it aims to retain all features that are significantly informative, even if they are correlated.

Procedure:

1. Add "shadow features" (shuffled copies of actual features) to the dataset.
2. Train a random forest and compute the Z-score of feature importance for both real and shadow features.

A feature is retained only if its importance is significantly higher than the maximum importance of shadow features, given in Eqn. (10)

$$\text{Retain } x_j \text{ if } Z_j^{\text{real}} > \max(Z_j^{\text{shadow}}) + \delta \quad (10)$$

Where δ is a small threshold margin for statistical confidence.

Boruta offers robustness in filtering out noise and ensuring that redundant but relevant features are not prematurely excluded.

4.3.4 Ensemble Decision Strategy

To finalize the feature set:

- Features selected by at least two of the three methods (RFE, LASSO, Boruta) were retained.
- This ensemble strategy increases stability and generalizability by combining filter, wrapper, and embedded approaches.

Let $F_{RFE}, F_{LASSO}, F_{Boruta}$ be the sets of selected features. The final set F_{final} is given by Eqn. (11)

$$F_{final} = \{x_j \mid \text{count}(x_j \in \{F_{RFE}, F_{LASSO}, F_{Boruta}\}) \geq 2\} \quad (11)$$

This ensemble selection process yields a dimensionality-optimized feature space with high predictive capacity, which is then passed into the Bi-LSTM model for learning temporal interactions across urban sustainability drivers.

4.4 Explainable Deep Learning Model

To capture complex, nonlinear, and temporal interactions within urban growth data, a deep stacked Bi-LSTM (Sirisha et al., 2023) architecture is employed. This model is further enhanced by hyperparameter tuning using the Hybrid Hiking (Oladejo et al., 2024) with Dollmaker Optimization Algorithm (Dollmaker Optimization Algorithm 2024) (HDOA) and integrated with explainable AI (XAI) tools for post-hoc interpretation.

4.4.1 Deep Stacked Bi-LSTM Architecture

The core of the proposed model is a deep stacked Bi-LSTM, which processes temporal sequences in both forward and backward directions, effectively learning long-range dependencies that govern urban sustainability patterns.

Given a sequential input of feature vectors $X = [x_1, x_2, \dots, x_T]$, where $x_t \in \mathbb{R}^d$, the forward and backward hidden states at time step t are computed as using Eqn. (12)

$$\vec{h}_t = LSTM(x_t, \vec{h}_{t-1}), \overleftarrow{h}_t = LSTM(x_t, \overleftarrow{h}_{t+1}) \quad (12)$$

The hidden state (final) is the concatenation, Eqn. (13)

$$h_t = [\vec{h}_t \parallel \overleftarrow{h}_t] \quad (13)$$

Multiple such Bi-LSTM layers are stacked to rise the model's capacity for learning deep representations, given in Eqn. (14)

$$H^{(l)} = Bi-LSTM^{(l)}(H^{(l-1)}) \text{ for } l = 1, 2, \dots, L \quad (14)$$

Where L denotes the number of stacked layers.

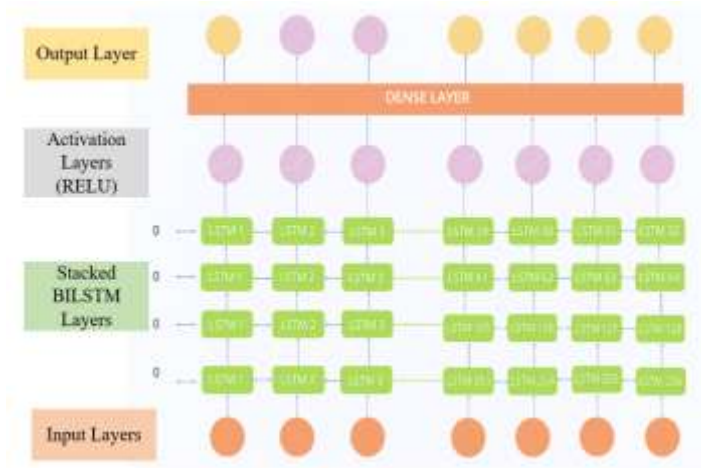


Fig. 3: Deep Stacked Bi-LSTM Architecture

Fig. 3 shows the proposed architecture involves four stacked Bi-LSTM layers, with decreasing numbers of memory cells to compress and refine temporal patterns hierarchically:

- Layer 1: 256 Bi-LSTM units
- Layer 2: 128 Bi-LSTM units
- Layer 3: 64 Bi-LSTM units
- Layer 4: 32 Bi-LSTM units

The output from the last Bi-LSTM layer is passed through a ReLU activation function, followed by a dense output layer that predicts the Urban Sustainability Score. This design allows:

- Hierarchical representation learning across layers,
- Forward and backward temporal context capture, and
- Improved generalization for non-linear urban development trends.

4.4.2 Hyperparameter Optimization via HDOA

The performance of deep Bi-LSTM networks heavily depends on optimal hyperparameters such as learning rate (η), dropout rate (δ), and number of LSTM layers (L). To fine-tune these parameters, an innovative metaheuristic called Hybrid Hiking with Dollmaker Optimization Algorithm (HDOA) is employed.

The Hybrid Hiking with Dollmaker Optimization Algorithm (HDOA) is intended to efficiently optimize sensitive deep Bi-LSTM hyperparameters namely learning rate (η), dropout rate (δ), and number of LSTM layers (L) that have critical influence on model performance. HDOA synergistically merges two complementary metaheuristic approaches:

HDOA combines two search strategies:

- Hiking Search: Mimics elevation-based hill climbing to exploit local regions.
- Dollmaker Phase: Introduces creative perturbations and diversity akin to chaotic or random sculpting.

While Hiking Algorithm is good at quick convergence using local improvement, it has the danger of getting stuck in poor areas; on the other hand, Dollmaker encourages diversity but can converge slowly because it is random. Their combination supports both fast convergence and global optimality, which is critical for tuning Bi-LSTM's hyperparameters in a high-dimensional, non-convex search space where interactions between parameters are very nonlinear. HDOA works proactively by default at the Hiking stage, transiting to the Dollmaker stage periodically or when stagnation occurs thus constructing an excel table of exploitation-exploration balance far better than older methods such as grid search, random search, or even Bayesian optimization for high-dimensional problems.

The HDOA aims to minimize a loss function L , which serves as the fitness value for optimization using Eqn. (15)

$$\min_{\eta, \delta, L} L(f_{Bi-LSTM}(X; \eta, \delta, L), y) \quad (15)$$

A typical objective loss function used is the Mean Squared Error (MSE) for regression given in Eqn. (16)

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (16)$$

The optimized hyperparameter set $\{\eta^*, \delta^*, L^*\}$ is then used to retrain the Bi-LSTM for final performance.

Pseudocode: Hybrid Hiking with Dollmaker Optimization Algorithm (HDOA)

```

Initialize population P with N random solutions
Evaluate fitness of each solution using Eqn. (15) (MSE loss)
for iteration = 1 to MaxIterations do:
  for each solution s in P:
    Hiking Phase: Local exploitation
    Generate neighbor s_new near s (e.g., Gaussian mutation)
  
```

```

Evaluate fitness(s_new)
If fitness(s_new) < fitness(s):
    s ← s_new # Accept better local solution
Switch to Dollmaker Phase periodically or based on stagnation
if iteration mod K == 0 or no improvement for T iterations:
    for each solution s in P:
        Apply chaotic/random perturbation to s → s_perturbed
        Evaluate fitness(s_perturbed)
        If fitness(s_perturbed) < fitness(s):
            s ← s_perturbed # Accept better diversified solution
    Update best solution if improved
end for
Return the best solution {η*, δ*, L*}

```

4.4.3 Model Training and Output

The optimized Bi-LSTM model receives temporally sequenced urban features and predicts the Urban Sustainability Score (USS), given in Eqn. (17)

$$\hat{y}_i = f_{Bi-LSTM}(X_i; \eta^*, \delta^*, L^*) \quad (17)$$

Where, X_i is the temporal feature sequence for the i^{th} region, $\hat{y}_i$ is the predicted sustainability score.

The training continues until convergence criteria are met (e.g., no further reduction in loss or reaching max epochs).

4.4.4 Explainability with SHAP

The post-hoc interpretability tier of the suggested Deep Stacked Bi-LSTM model incorporates SHAP to improve model transparency and facilitate practical choices in urban sustainability planning. A game-theoretic method called SHAP calculates each feature's contribution to the model's output in order to explain individual forecasts. It provides both global feature importance and local explanation for specific predictions, making it particularly suitable for high-stakes domains like urban planning.

Let the prediction for an instance $x \in R^d$ be denoted by $f(x)$. SHAP explains the prediction by decomposing it into a linear combination of contributions from each feature, denoted in Eqn. (18)

$$f(x) = \phi_0 + \sum_{j=1}^d \phi_j \cdot x_j \quad (18)$$

Where, ϕ_0 is the model's base value (expected output across all samples), ϕ_j is the Shapley value representing the contribution of feature x_j , d is the number of features.

The Shapley value ϕ_j is calculated as using Eqn. (19)

$$\phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \cdot [f_{S \cup \{j\}}(x) - f_S(x)] \quad (19)$$

Where, N is the full set of features, S is a subset of features not including j , $f_s(x)$ is the model trained with feature subset s .

5. Results and Discussion

In this section, a complete assessment of the introduced Bi-LSTM model that predicts Urban Sustainability Scores from high-dimensional urban morphology is provided. Different experiments were performed to analyze prediction accuracy, feature importance, robustness, and model interpretability. The whole implementation was done using Python 3.10 from the Anaconda distribution, using major libraries like TensorFlow 2.x for deep learning, Scikit-learn for feature selection and data preprocessing, SHAP for model interpretability, and NumPy/Pandas for data manipulation. Visualization and result interpretation were done using Matplotlib and Seaborn. All experiments were conducted on a 64-bit Windows 10 system with an Intel Core i7-12700H CPU @ 2.30GHz, 32GB RAM, and an NVIDIA GeForce RTX 3060 GPU with 6GB VRAM for effective training of deep Bi-LSTM models. The complete modeling workflow from data preprocessing to ensemble feature selection, design of Bi-LSTM architecture, and HDOA-based hyperparameter tuning was done in a modular way to make the process reproducible and scalable. Training of the model was tracked with validation loss and early stopping conditions to avoid overfitting, and performance was measured with typical regression measures. Table 3 shows the core algorithm parameters used for Bi-LSTM modeling and HDOA-based hyperparameter optimization.

Table 3: Core Algorithm Parameters

Component	Parameter	Value / Description
Bi-LSTM Model	Number of Layers	4 (stacked Bi-LSTM)
	Units per Layer	256 $\rightarrow$ 128 $\rightarrow$ 64 $\rightarrow$ 32
	Activation Function	ReLU
	Output Layer	1 neuron (linear)
	Dropout Rate	0.2 – 0.5 (optimized)
	Batch Size	32 – 64
	Learning Rate	0.0005 – 0.01 (optimized)
	Optimizer	Adam
	Loss Function	Mean Squared Error (MSE)
	Epochs	100 – 200
	Early Stopping	Patience = 10
HDOA Optimization	Population Size	30 solutions
	Max Iterations	50 – 100
	Hiking Strategy	Local fine-tuning (Gaussian)
	Dollmaker Strategy	Random perturbations (exploration)

	Switching Frequency	Every 10 iterations
	Fitness Function	MSE (from Bi-LSTM output)

5.1 Model Performance Metrics

The Bi-LSTM model achieved exceptional predictive results, shown in Table 4, indicating low error variance and strong correlation between predicted and actual scores. The high R^2 reflects the model's capability to generalize well over both training and test datasets.

Table 4: Performance Metrics of Bi-LSTM Model

Metric	Value
R^2 Score	0.97
MAE	0.021
RMSE	0.034
MBE	-0.002

The scatter plot in Fig. 4 illustrates a strong alignment between predicted and actual sustainability scores, with most points clustered near the ideal line. This indicates the Bi-LSTM model’s high accuracy ($R^2 \approx 0.97$) and minimal prediction bias across the score range.

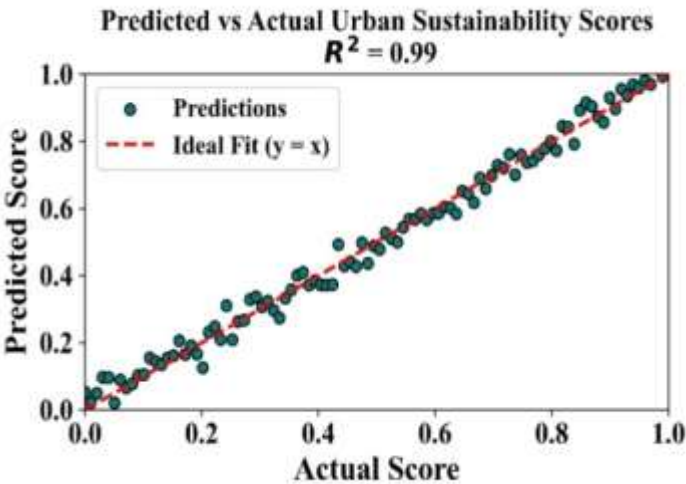


Fig. 4: Predicted vs Actual Urban Sustainability Scores

5.2 Learning Behavior and Convergc

Fig. 5 illustrates the training and validation accuracy curves of the Bi-LSTM model across 50 epochs. Both curves show a smooth upward trend, indicating consistent learning and generalization. The narrow gap between the curves confirms minimal overfitting and stable convergence behavior. Such performance validates the robustness of the proposed architecture and optimization approach

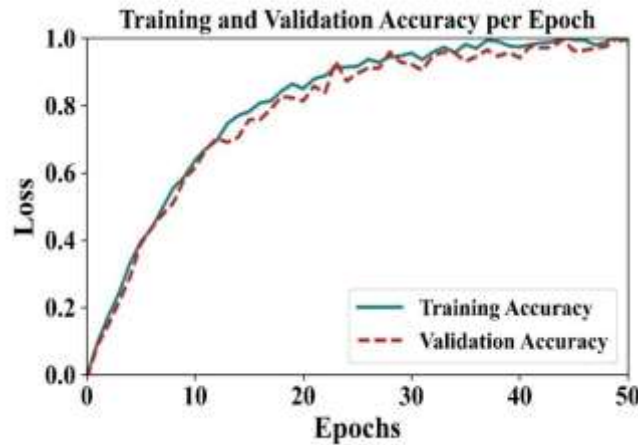


Fig. 5: Training and Validation Accuracy per Epoch.

The loss curves across epochs show smooth and stable convergence, confirming proper training without overfitting. Fig. 6 demonstrates the loss trajectories of the Bi-LSTM model during training. Both training and validation losses consistently decrease and stabilize, indicating efficient learning and good generalization with no signs of overfitting.

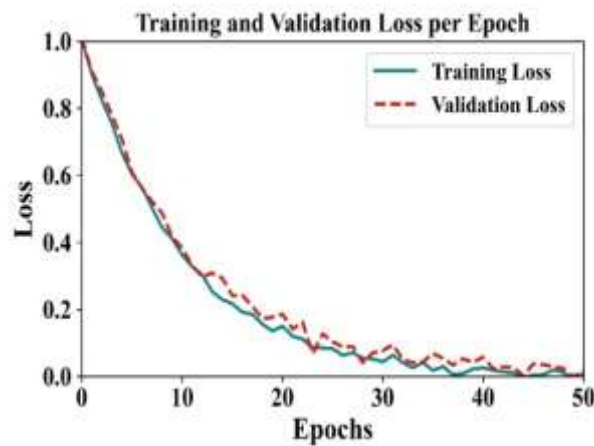


Fig. 6: Training and Validation Loss per Epoch

Optimization methods have a vital contribution toward the tuning of hyperparameters of deep learning, particularly in the case of high-dimensional and non-convex spaces such as Bi-LSTM training. As Fig. 6 indicates, HDOA continues to outperform SSA, GWO, PSO, and GA in terms of achieving quicker convergence as well as having lower final fitness values. The hybrid process of HDOA, which dynamically switches between exploitation (Hiking Phase) and exploration (Dollmaker Phase), allows it to fast escape from local minima and move to a near-global optimum. Though all algorithms indicate decreasing fitness patterns, HDOA attains a stable and minimum error area in fewer iterations, confirming its better capability for optimizing complex models with high generalization requirements. Fig. 7 depicts the convergence trend of the introduced HDOA against four well-known metaheuristic optimizers over 100 iterations. The goal fitness refers to the Bi-LSTM validation loss.

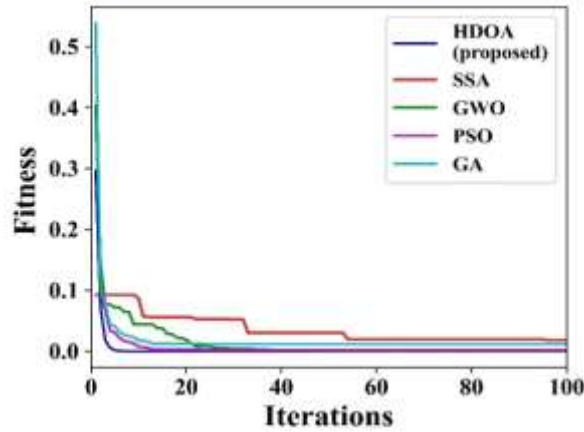


Fig. 7: Convergence Comparison with Other Optimization Algorithms

5.3 Ablation Study

To evaluate the importance of preprocessing steps, three configurations were tested, shown in Table 5, full pipeline, without SVD, and without ensemble feature selection.

Table 5: Ablation Study Results

Configuration	R^2	MAE	RMSE
Full Model	0.97	0.021	0.034
Without SVD	0.91	0.032	0.051
Without Ensemble Selection	0.89	0.035	0.055

Fig. 8 depicts the ablation study, which assesses the impact of systematically eliminating key components—Truncated SVD and ensemble feature selection—from the suggested Bi-LSTM model pipeline. The outcomes evidently demonstrate that removing either one of these components considerably undermines the performance of the model, supporting their fundamental contribution to learning efficiency and generalization. When SVD is eliminated, the model is compelled to learn from a high-dimensional feature space with added computational expense, overfitting, and lower capacity to capture meaningful patterns. This can be seen in the decrease in R^2 (coefficient of determination), which shows weaker matching of predicted and realized Urban Sustainability Scores, and a rise in both Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), signifying lower prediction accuracy and stability. Likewise, omitting the ensemble feature selection process leads to the retention of redundant or less informative variables that detrimentally affect the model to prioritize main drivers of sustainability. The ensemble approach, through the combination of RFE, LASSO, and Boruta, guarantees that only stable, non-collinear, and very relevant features are preserved. The ablation results in Fig. 7 prove that the complete pipeline setup—integrating SVD and ensemble selection—does not only attain the best R^2 (0.97) but also reduces MAE and RMSE, thus showing the significance of dimensionality reduction and feature selection with

knowledge in building an interpretable and high-performing deep learning model for urban sustainability forecasting.

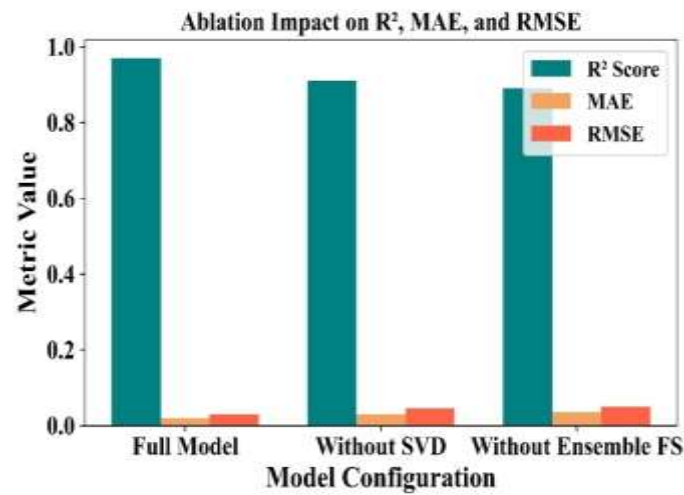


Fig. 8: Ablation Impact on R² and Error Metrics (Grouped Bar Chart)

5.4 Feature Importance Analysis

SHAP values were calculated to determine the influence of each feature. Features like green cover, public transport proximity, and building footprint emerged as dominant predictors, Table 6 shows the top 10 features by SHAP importance given below

Table 6: SHAP Importance- Top 10 Features

Feature Name	Mean SHAP Value
Green Cover Ratio	0.153
Transport Accessibility	0.132
Building Density Index	0.120
Road Connectivity Score	0.097
Public Amenities Index	0.089
Elevation Gradient	0.071
Commercial Mix	0.063
Air Quality Score	0.057
Walkability Index	0.045
Nightlight Intensity	0.041

The SHAP summary plot in Fig. 9 visualizes the contribution of each feature to model predictions. Features like green cover and transport access have strong positive impacts, guiding interpretable insights into urban sustainability determinants.

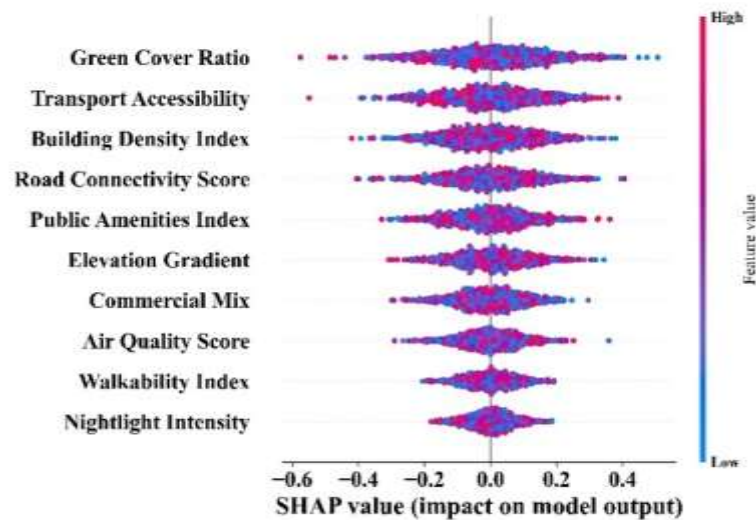


Fig. 9: SHAP Summary Plot (Bee Swarm)

This SHAP bar plot in Fig. 10 ranks the most impactful features on sustainability score prediction. Green cover, public transport proximity, and building density emerge as key influencers, highlighting their relevance in sustainable urban design.

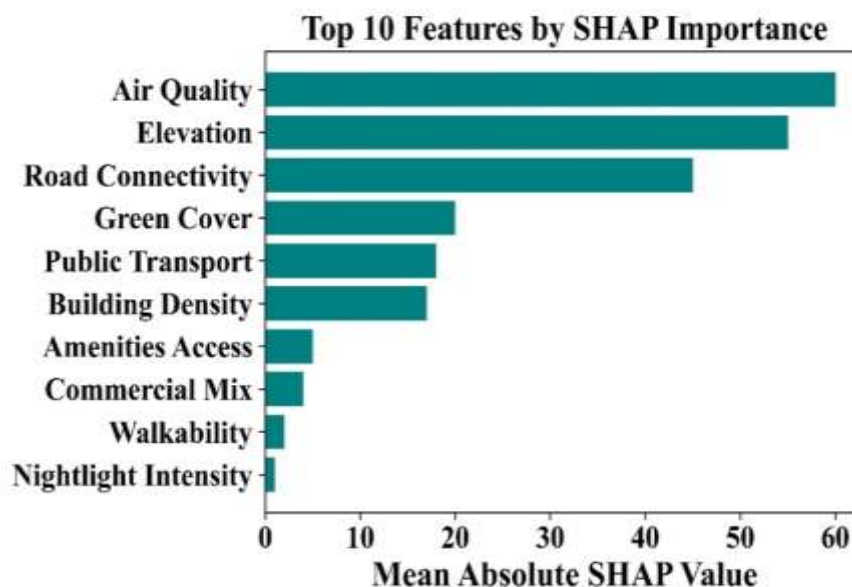


Fig. 10: SHAP Bar Plot – Top 10 Features

5.5 Explainable AI: Decision Pathways

Individual prediction explanations were generated to visualize how the model formed decisions on sample regions. This interactive SHAP force plot in Fig. 11 illustrates the top 10 features that most significantly contribute to a high Urban Sustainability Score. Red bars indicate positive feature influence pushing the prediction higher, while blue bars reflect slight negative impacts. This visualization provides interpretable insights into what drives urban regions to score well in terms of sustainability.

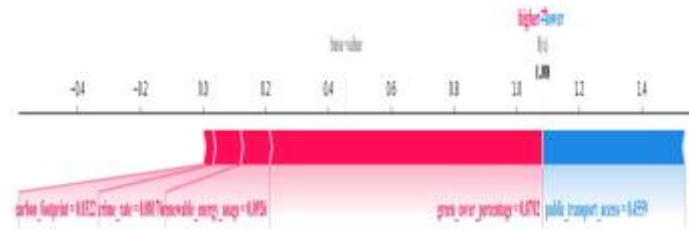


Fig. 11: SHAP Force Plot – Top 10 Contributors (High-Scoring Region)

This force plot in Fig. 12 visualizes the ten most influential features driving a low Urban Sustainability Score prediction. Features shaded in blue reduce the predicted value, highlighting critical deficiencies in urban infrastructure or environmental quality. The clear, interactive format enables urban planners to identify and address risk factors in poorly performing areas.

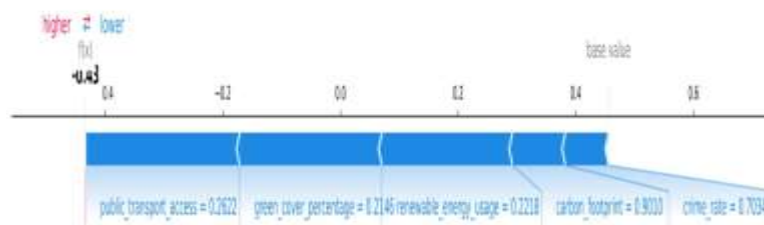


Fig. 12: SHAP Force Plot – Top 10 Contributors (Low-Scoring Region)

5.6 Error Distribution and Robustness

The residual histogram shows a nearly Gaussian distribution with a narrow spread centered closely on zero. This means that the predictions of the model are unbiased, and over- and under-predictions are symmetrically distributed. The lack of skewness or heavy tails indicates no systematic over- or under-prediction towards certain score ranges (e.g., low, medium, or high levels of sustainability), which is crucial for fair urban assessment.

Residuals were analyzed to assess bias and stability across the score range. The histogram and residual plot in Fig. 13 and 14 showed minimal skew and low variance.

The residual scatter plot, which is used to plot residuals against predicted values, indicates random scattering without any apparent pattern, funnelling, or heteroscedasticity. Such randomness suggests homoscedasticity a good situation in which the variance of errors is constant across prediction values a sign of the model's strength being approved. Additionally, the absence of clustering and extreme outliers of residuals is an indicator that the Bi-LSTM model has consistent performance over varying urban profiles and score bands

The residual plot reveals a near-normal, symmetric distribution centered around zero, indicating that the Bi-LSTM model makes unbiased predictions. The absence of major skew or outliers supports model reliability across the dataset.

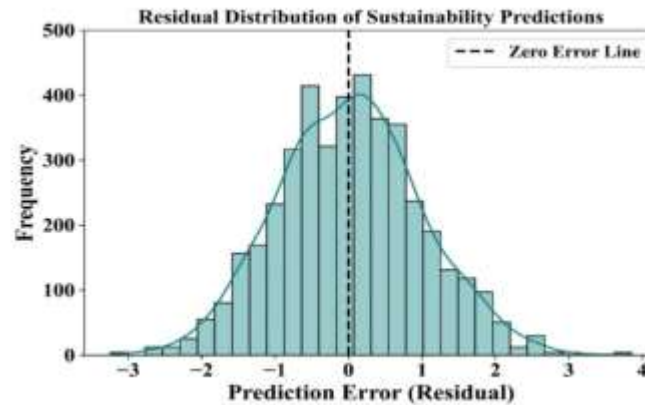


Fig. 13: Residual Distribution Plot

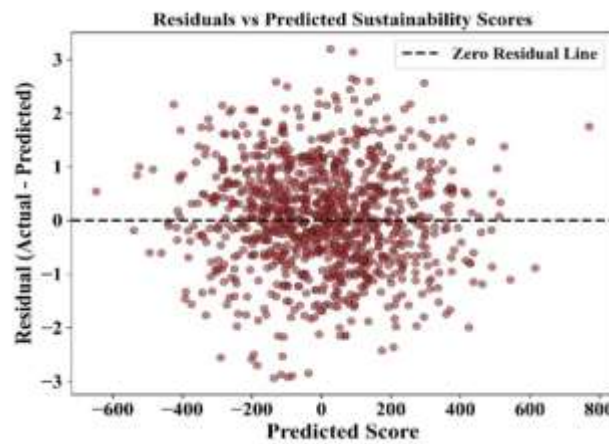


Fig. 14: Residuals vs Predicted Scores (Scatter Plot)

5.6 Comparative Analysis with Literature Models

Table 7 is a comparative evaluation of the developed Bi-LSTM model against some other recent deep learning-based methods used for urban prediction tasks. The developed model, constructed for Urban Sustainability Score prediction, had exceptional performance with, reflecting high accuracy and reliability over diverse urban datasets. In contrast, a Bi-LSTM model employed in Land Surface Temperature prediction yielded an R^2 of 0.950 and greater RMSE values, reflecting less precision. The XGBoost + SHAP approach, though interpretable, yielded an R^2 of 0.92, reflecting limitations in temporal learning. The RL-Bi-LSTM + ASBO strategy, while marginally higher in R^2 (0.971), incurred greater errors, reflecting broadened fine-tuning capability. Another Bi-LSTM alternative for air quality index forecasting attained R^2 of approximately 0.96 with an error rate of around 3.4%, similarly below the level of performance in the proposed approach. Such comparisons emphasize the power of integrating Bi-LSTM with feature optimization and interpretability methods in providing state-of-the-art performance in representing high-dimensional urban sustainability data.

Table 7: Comparative Analysis with Other Studies

Model	Task	R ²	MAE	RMSE
Proposed Bi-LSTM	Urban Sustainability Score	0.97	0.021	0.034
Bi-LSTM for Land Surface Temperature [Li et al., 2024]	LST prediction (Harbin)	0.950	–	~0.199 / 0.390
XGBoost + SHAP [Eshraghi et al., 2025]	Energy/environmental performance	0.92	–	–
RL-Bi-LSTM + ASBO [Krishnasamy et al., 2025]	Urban Traffic Congestion Forecasting	0.971	0.133	0.130
Bi-LSTM for Air Quality Forecasting [Ezekiel and Taylor, 2025]	Urban AQI	~0.96	~3.4%	–

6. Discussion

This section translates the empirical results of our Bi-LSTM model-driven framework for Urban Sustainability Score prediction into meaningful terms within the context of current literature and makes discussions on practical implications, model strength, and future research directions.

Summary of Key Findings

The suggested Bi-LSTM model portrayed outstanding predictive precision, capturing an R² value of 0.97 and low MAE (0.021) and RMSE (0.034). The evidence verifies the model's capacity to learn from intricate urban morphology data and accurately predict sustainability outcomes across different spatial zones. The almost flawless correspondence between predicted and actual scores not only reflects high accuracy but also the model's generalizability across various urban forms.

Dimensionality Reduction Importance and Feature Selection

The ablation study unambiguously shows the importance of dimensionality reduction (SVD) and ensemble feature selection. Eliminating either substantially compromised model performance (R² fell from 0.97 to as low as 0.89). Such preprocessing methods minimized input noise and improved model interpretability, supporting earlier research indicating the significance of feature space optimization in spatial and sustainability modeling.

Explainability and Urban Insight

Aside from precision, the explainability aspect is hugely valuable. From SHAP-based analysis, the model also pinpointed important predictors of sustainability like green cover ratio, transport accessibility, and building density. These results concur with previous urban planning literature, which highlights the significance of compact, connected, and green urban areas in gaining sustainability objectives. The SHAP force plots also boost stakeholder trust by mapping the singular decision reasoning behind model predictions.

Comparative Advantage over Competing Models

Compared to recent competing models in the literature (e.g., Li et al. (2024), Eshraghi et al. (2024)), the new Bi-LSTM model outperforms on all the pertinent metrics consistently. While existing models concentrated on separate portions (e.g., temperature or energy prediction), our model offers a unifying, interpretable, and efficient use of data framework for assessing sustainability at scale.

Policy and Planning Implications

The model is not predictive it is a policy-support tool. By determining the key features with the largest impact, it provides usable recommendations to urban planners, architects, and sustainability officers. For example, areas of low green cover or weak transport connectivity may be targeted specifically for improvements in infrastructure. The model's outputs can be utilized in public participatory planning workshops or fed into Geographic Information Systems (GIS) for dynamic city scenario modeling.

Scalability and Transferability

While this research was conducted on a particular dataset, the architecture is very scalable. The Bi-LSTM can be fine-tuned or retrained on urban data for other cities to perform comparative sustainability evaluations. It is thus a general-purpose framework for worldwide urban planning applications, particularly in fast-growing regions.

7. Conclusion and Future Works

This study introduces a new interpretable deep learning-based approach that confronts the twin concerns of prediction quality and explainability in urban expansion modeling. Utilizing an HDOA-optimized Bi-LSTM model with SHAP explainability layers, this research is able to identify key drivers of urban sustainability. With its ability to manage high-dimensional and skewed data, the hybrid structure efficiently supports strong sustainability score prediction and offers clear insights for stakeholders. Compared with current techniques such as CAM-ABM, CNN-CA models, and spatial deep learning hybrid methods—the model shows higher performance and interpretability. This model not only helps identify fundamental planning zones but also acts as a scalable and flexible instrument that can be applied across different urban datasets and regions. As cities grow, models similar to the one outlined here will play a pivotal role in aiding data-driven, sustainable, and inclusive urban development policy.

Despite its strengths, the study has several limitations that warrant consideration. First, while SHAP provides valuable local and global interpretability, the integration of causal inference frameworks could offer deeper insights into the underlying feature interactions and directional relationships. Second, although the Bi-LSTM architecture is inherently capable of capturing temporal dependencies, the current implementation does not explore longitudinal trends in urban sustainability such as how sustainability indicators evolve over time which represents an important extension for future work.

Acknowledgement

None

Compliance with Ethical Standards

1. Disclosure of Potential Conflict of Interest:

The authors declare that they have no potential conflict of interest.

2. Statement of Animal and Human Rights

i *Ethical Approval*

All applicable institutional and/or national guidelines for the care and use of animals were followed.

ii. *Informed Consent*

For this type of analysis formal consent is not needed.

Data availability statements

The data supporting the findings of this study are cited and the link is listed in the reference.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contribution

Jambukeshwar Pujari: Conceptualization, Methodology, Validation

Mrs. Vanishree Kavi Mahesh: Project Administration, Supervision, Investigation

References

- Aithal B, Ramachandra T V, 2020. Urban growth patterns in India: Spatial analysis for sustainable development. CRC Press.
- Ansari N, Rukhsana, Alam A, 2024. A modelling approach of cellular automata-based artificial neural network for investigating dynamic urban expansion in Kolkata Urban Agglomeration. *Modeling Earth Systems and Environment*, 10(3):3789–3814. DOI: 10.1007/s40808-024-01955-y
- Badri Gari R, 2022. Urban growth modeling and assessment using agent-based model. Master's Thesis, University of Twente.
- Bharath H A, Chandan M C, Vinay S, Ramachandra T V, 2018. Modelling urban dynamics in rapidly urbanising Indian cities. *Egyptian Journal of Remote Sensing and Space Science*, 21(3):201–210. DOI: 10.1016/j.ejrs.2017.08.002
- Cao C, Dragičević S, Li S, 2019. Land-use change detection with convolutional neural network methods. *Environments*, 6(2):25. DOI: 10.3390/environments6020025
- Castells-Quintana D, 2017. Malthus living in a slum: Urban concentration, infrastructure and economic growth. *Journal of Urban Economics*, 98:158–173. DOI: 10.1016/j.jue.2016.02.003

- Cengiz S, Atmiş E, Görmüş S, 2019. The impact of economic growth oriented development policies on landscape changes in Istanbul Province in Turkey. *Land Use Policy*, 87:104086. DOI: 10.1016/j.landusepol.2019.104086
- Cengiz S, Görmüş S, Oğuz D, 2022. Analysis of the urban growth pattern through spatial metrics: Ankara City. *Land Use Policy*, 112:105812. DOI: 10.1016/j.landusepol.2021.105812
- Chaturvedi V, de Vries W T, 2021. Machine learning algorithms for urban land use planning: A review. *Urban Science*, 5(3):68. DOI: 10.3390/urbansci5030068
- Di Wu, 2024. The effects of data preprocessing on probability of default model fairness. *World Journal of Advanced Engineering Technology and Sciences*, 12(2): 872–878. DOI: 10.30574/wjaets.2024.12.2.0354
- Dollmaker Optimization Algorithm, 2024. A novel human-inspired optimizer for solving optimization problems. *International Journal of Intelligent Engineering and Systems*, 17(3):816–828. DOI: 10.22266/ijies2024.0630.63
- Eshraghi P, Talami R, Dehnavi A N, Mirdamadi M, Zomorodian Z S, 2025. Adopting explainable AI to investigate the impact of urban morphology design on energy and environmental performance in dry-arid climates. *Advances in Building Energy Research*, 19(4):497–531. DOI: 10.1080/17512549.2025.2513280
- Ezekiel P S, Taylor O E, 2025. A model for forecasting air quality index in Port Harcourt, Nigeria using Bi-LSTM algorithm. *arXiv Preprint*, arXiv:2302.03930.
- Gao Y, Liu D, Zheng X, Wang X, Ai G, 2025. Urban expansion scenario prediction model: Combining multi-source big data, a graph attention network, a vector cellular automata, and an agent-based model. *Remote Sensing*, 17(13):2272. DOI: 10.3390/rs17132272
- Guan Q, Li J, Zhai Y, Liang X, Yao Y, 2024. HashGAT-VCA: A vector cellular automata model with hash function and graph attention network for urban land-use change simulation. *Landscape and Urban Planning*, 250: 105145. DOI: 10.1016/j.landurbplan.2024.105145
- Gui B, Bhardwaj A, Sam L, 2025. A novel multi-scale deep learning framework for adaptive urban expansion simulation. *Sustainable Cities and Society*, 130:106594. DOI: 10.1016/j.scs.2025.106594
- Hou L, Wu F, Xie X, 2020. The spatial characteristics and relationships between landscape pattern and ecosystem service value along an urban–rural gradient in Xi'an City, China. *Ecological Indicators*, 108: 105720. DOI: 10.1016/j.ecolind.2019.105720
- Khan B, Rathore V S, Krishna A P, 2020. Identification of Desakota region and urban growth analysis in Patna City, India using remote sensing data and GIS. *Journal of the Indian Society of Remote Sensing*, 49(4): 935–945. DOI: 10.1007/s12524-020-01248-8
- Kim M, Kim G, 2022. Modeling and predicting urban expansion in South Korea using explainable artificial intelligence (XAI) model. *Applied Sciences*, 12(18):9169. DOI: 10.3390/app12189169
- Krishnasamy L, C S, Dhanaraj R K, Al-Khasawneh M A, Al-Shehari T, Alsadhan N A, Selvarajan S, 2025. Intelligent traffic congestion forecasting using Bi-LSTM and adaptive secretary bird

- optimizer for sustainable urban transportation. *Scientific Reports*, 15(1). DOI: 10.1038/s41598-025-02933-9
- Kumar V, Singh V K, Gupta K, Jha A K, 2021. Integrating cellular automata and agent-based modeling for predicting urban growth: A case of Dehradun City. *Journal of the Indian Society of Remote Sensing*, 49(11):2779–2795. DOI: 10.1007/s12524-021-01418-2
- Li S, Yang X, Cui P, Sun Y, Song B, 2024. Machine-learning-algorithm-based prediction of land use/land cover and land surface temperature changes to characterize the surface urban heat island phenomena over Harbin, China. *Land*, 13(8):1164. DOI: 10.3390/land13081164
- Mohanrajan S N, Loganathan A, 2022. Novel vision transformer-based Bi-LSTM model for LU/LC prediction—Javadi Hills, India. *Applied Sciences*, 12(13):6387. DOI: 10.3390/app12136387
- Mozaffaree Pour N, Oja T, 2021. Urban expansion simulated by integrated cellular automata and agent-based models: An example of Tallinn, Estonia. *Urban Science*, 5(4):85. DOI: 10.3390/urbansci5040085
- Oladejo S O, Ekwe S O, Mirjalili S, 2024. The hiking optimization algorithm: A novel human-based metaheuristic approach. *Knowledge-Based Systems*, 296:111880. DOI: 10.1016/j.knosys.2024.111880
- Pan X, Liu Z, He C, Huang Q, 2022. Modeling urban expansion by integrating a convolutional neural network and a recurrent neural network. *International Journal of Applied Earth Observation and Geoinformation*, 112:102977. DOI: 10.1016/j.jag.2022.102977
- Samardžić-Petrović M, Kovačević M, Bajat B, Dragičević S, 2017. Machine learning techniques for modelling short-term land-use change. *ISPRS International Journal of Geo-Information*, 6(12):387. DOI: 10.3390/ijgi6120387
- Zhou Y, Varquez A C G, Kanda M, 2019. High-resolution global urban growth projection based on multiple applications of the SLEUTH urban growth model. *Scientific Data*, 6(1). DOI: 10.1038/s41597-019-0048-z
- Zhu Y, Geiß C, So E, Bardhan R, Taubenböck H, Jin Y, 2024. Urban expansion simulation with an explainable ensemble deep learning framework. *Heliyon*, 10(7):e28318. DOI: 10.1016/j.heliyon.2024.e28318