CINEFORUM
ISSN: 0009-7039
Vol. 66. No. S6, 2026 (Special Issue 6)

Integrated Remote Sensing and Machine Learning Assessment of Environmental Degradation in Haryana (2000–2023).

Sonu Tak[1], Manoj Kumar[2]

[1] Research Scholar, Department of Economics, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India.
Email: jakharsonu01@gmail.com
[2] Assistant Professor, Department of Economics, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India.
Email: manojkumareconomics@gmail.com

Abstract
Environmental degradation has been exacerbated in fast growing urban and agricultural areas, and integrated assessment frameworks are needed to facilitate sound policy action. This study has been carried out in the state of Haryana by adopting remote sensing, statistical and machine learning techniques to assess the environmental degradation in the state for the period 2000-2023. The composite Environmental Degradation Index (EDI) was created through Principal Component Analysis (PCA) incorporating various indices like, Normalised Difference Vegetation Index (NDVI), Land Surface Temperature (LST), Land Use/Land Cover (LULC), groundwater levels and air quality parameters.
The Mann–Kendall trend analysis and Sen's slope estimator indicated that the NDVI and groundwater levels had significantly decreased while the temperature LST and particulate pollution had increased. The regression analysis showed that urbanisation and agricultural intensity were the two main factors that accounted for a significant amount of variability ($R^2 = 0.78$) in environmental degradation. In comparison to other predictive models, the model Random Forest showed better performance ($R^2 = 0.91$), which is higher than the models Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM) that were able to capture the complex non-linear relationships.
The geographical analysis revealed high degradation zones in NCR districts. Projections based on scenarios suggest an increase of 12–18% in environmental degradation, but a decrease of up to 10% is possible with targeted policy interventions. The study provides a sound, scientific basis for sustainable land-use planning, groundwater management and pollution control measures.....
Keywords: Environmental Degradation Index (EDI); Remote Sensing; Machine Learning; Random Forest; Land Surface Temperature (LST); Land Use Land Cover (LULC); Spatiotemporal Analysis

Nomenclature

Symbol	Description
NDVI	Normalised Difference Vegetation Index
LST	Land Surface Temperature
LULC	Land Use/Land Cover
EDI	Environmental Degradation Index
PCA	Principal Component Analysis
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
R^2	Coefficient of Determination.

INTRODUCTION

In the areas with high rates of urbanisation and intensive agriculture, environmental degradation is a burning issue. These processes cause great pressure on the natural resources resulting in decreasing vegetation cover, groundwater depletion, rising land surface temperatures (LST) and worsening air quality. These elements combined bring severe threats to the ecological balance, human health, and sustainability.
In India, these issues are especially noticeable because of the twin stresses of population increase and development that is resource intensive. Haryana is a major case; it is not only a major agricultural state but also a fast urbanising state in the National Capital Region (NCR). Although agricultural intensification has led to food security, it has also caused too much extraction of groundwater and degradation of soils [3,13]. Simultaneously, land-use change, rising surface temperatures, and air pollution have also risen in the cities of the Gurugram and Faridabad districts due to urban expansion [6,7]. Also, agricultural activities in the area like stubble burning, in conjunctiṇ have led to long-term air quality degradation in the area [7,18].

Remote sensing indicators (Normalised Difference Vegetation Index (NDVI) and land surface temperature (LST)) offer useful data about the state of the environment and have been extensively utilized in long-term monitoring. A decrease in NDVI and an increase in LST are generally attributed to vegetation loss and urban heat island effects [20]. The other serious indicator of the environmental stress is groundwater depletion especially in areas where agricultural activities are of high intensity and irrigation is a main requirement [3,19]. More so, escalating levels of particulate matter (PM2.5 and PM10) and the gaseous pollutants have been associated with environmental deterioration and health hazards to the population [7,18].

Although there is a lot of research available on the topics of individual environmental indicators, the majority of studies are conducted in a fragmented manner, as they study the groundwater, land-use change, or air pollution separately. These methods are not able to reflect the multi-dimensional and interrelated nature of environmental systems. Despite the fact that, composite indices have been suggested as a way of combining a number of indicators into a single framework [22], there is little use of composite indicators at the state level in India. Moreover, although statistical methods, including the Mann Kendall test, the slope estimator used by Sen, and so on, are popular in trend analysis [8,15], they only describe past trends, and not predict them.

Recent developments in machine learning provide better options to model non-linear and complex environmental systems. Random Forest, Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) networks are algorithms that have shown excellent performance in environmental prediction tasks [2,4,9,24]. The methods allow combining multi-source data sets and increase predictive accuracy. Nevertheless, their joint use along with remote sensing and statistical analysis in an integrated system of evaluating the environment on the state level is an unexplored area in the Indian environment.

To fill these gaps, this paper formulates a composite framework to evaluate environmental degradation in Haryana during 2000-2023. Principal Component Analysis (PCA) is used to construct a composite Environmental Degradation Index (EDI), based on multiple indicators such as NDVI, land surface temperature (LST), land use/land cover (LULC), groundwater levels, air quality parameters, climatic variables, and socio-demographic factors. The trend analysis is done through the Mann Kendall test and Sen slope estimator and predictive modelling is done through the use of Random Forest, Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM) models.

Moreover, analysis is conducted using scenarios to analyze the future environmental conditions in terms of business-as-usual, climate variability and policy-intervention scenarios. This method helps to find out the major factors contributing to the deterioration of the environment and gives an understanding of what could happen in the future.

The following are three main contributions of this study. First, it offers an integrated assessment of the environmental degradation at state level by multi-indicator, spatially explicit analysis based on integrated datasets. Second, it integrates both statistical and machine learning methods to enhance both diagnostic and predictive analysis. Third, it provides scenario-driven insights that can facilitate environmental planning, resource management, and policy decision-making based on data in rapidly developing areas.

A novel study was made consisting of an entirely new framework that presents a comprehensive evaluation of Environmental Degradation. This will be used by combining Remote Sensing Indicators, Statistical analysis and Machine Learning all under one framework. This study, rather than merely focusing on individual environmental factors like existing studies, develops an Environmental Degradation Index (EDI) through Principal Component Analysis (PCA). The Index captures the multi-dimensional aspect of environmental degradation.

In addition, this study uniquely combines Random Forest, Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM), which are complex predictive models, to explain non-linear relationships better and improve forecasting accuracy. The originality is enhanced further by the use of scenario-based analysis (business-as-usual, climate variability and policy intervention) which join past, present and future. In essence, this approach, which is integrated and driven by data, complements an important research gap. Moreover, it turns out to be handy in creating a scalable framework for environmental assessment at the state level.

The results of this study have important implications for policymakers, environmental planners and government agencies. An integrated framework can support sustainable land-use planning, groundwater management, pollution control strategies and climate adaptation policies. The research improves forecasting ability with the adoption of machine learning models, hence effective and proactive decision-making.

It is essential for a rapidly urbanising and agriculturally intensive region that has high environmental pressure. The methodology can be utilised in designing real-life environmental monitoring systems. The environmental monitoring systems provide more useful targeted interventions. It can also enhance the long-term management of resources. Furthermore, it will improve sustainability outcomes.

1.1 Statement of Novelty

This study provides a novel and integrated framework for assessing environmental degradation in Haryana by combining remote sensing, statistical analysis, and machine learning approaches within a single analytical model. Unlike earlier studies that primarily focused on individual environmental indicators such as vegetation loss, groundwater depletion, land-use change, or air pollution in isolation, the present research develops a composite Environmental Degradation Index (EDI) using Principal Component Analysis (PCA) to capture the multidimensional nature of environmental degradation. The study further enhances novelty by integrating advanced predictive models including Random Forest, Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) for forecasting future environmental trends. Additionally, the use of scenario-based analysis (Business-as-Usual, Climate Variability, and Policy Intervention) provides a comprehensive understanding of past trends, present conditions, and future environmental risks. Therefore, this study contributes a scalable, data-driven, and interdisciplinary framework for environmental degradation assessment at the state level.

1.2 Industrial Relevance

The findings of this study have significant industrial and practical relevance for environmental governance, sustainable resource management, and policy formulation. The integrated framework can support government agencies, environmental monitoring authorities, urban planners, and industrial regulators in identifying high-risk environmental degradation zones and designing targeted intervention strategies. The predictive modelling approach can improve decision-making related to land-use planning, groundwater conservation, pollution control, and climate adaptation measures. In addition, the use of geospatial technologies and machine learning tools can assist industries and environmental institutions in developing efficient monitoring systems for sustainable infrastructure planning and ecological risk assessment. The methodology can also be replicated in other rapidly urbanizing and agriculturally intensive regions for evidence-based environmental management and long-term sustainability planning.

2. Materials and Methods

The methodology adopted in this study integrates remote sensing, statistical analysis, and machine learning techniques to assess environmental degradation in Haryana over the period 2000–2023. The approach involves the extraction of environmental indicators from multi-source datasets, construction of a composite Environmental Degradation Index (EDI), analysis of temporal and spatial trends, and prediction of future environmental conditions using machine learning models. The overall framework is designed to capture both the current state and future trajectory of environmental degradation in a unified analytical structure.

2.1 Study area

Figure 1 shows the geographical location of the study area, Haryana, along with its district boundaries used for spatial analysis.

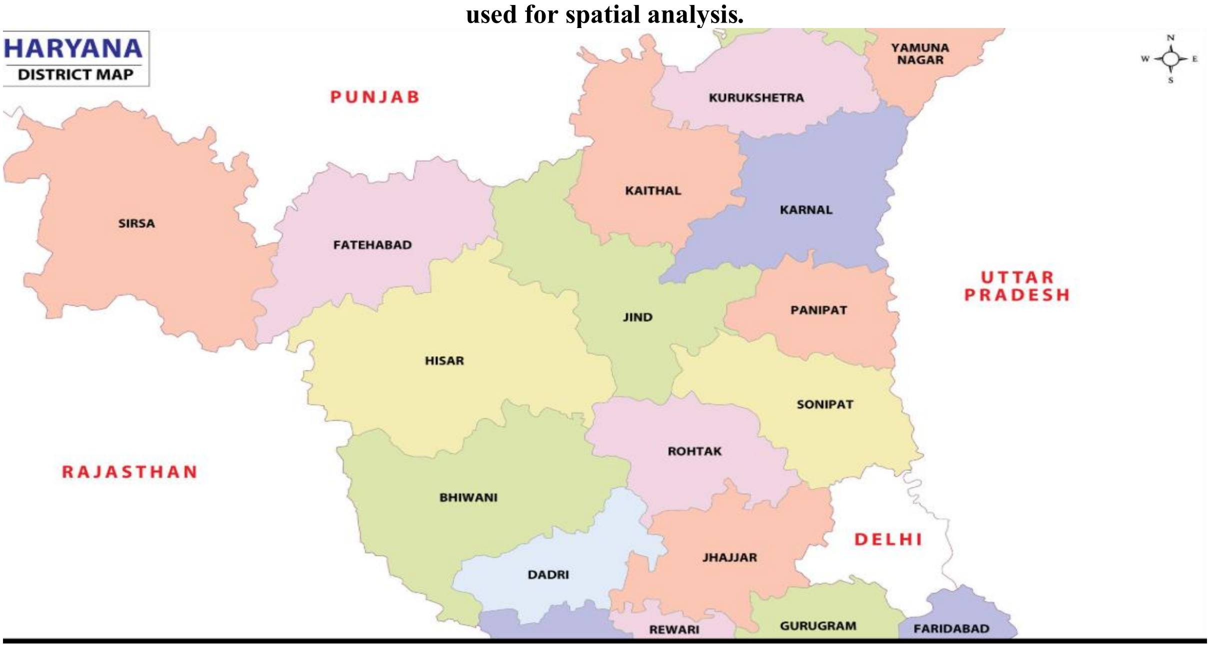

Figure 1: Geographical Location

As shown in Figure 1, Haryana's spatial diversity and proximity to the National Capital Region (NCR) make it a suitable case for analysing environmental degradation patterns. Haryana, a state in northwest India is located at 27°39′N to 30°55′N and 74°28′E to 77°36′E with an overall area of about426the km [43]. The state shares its borders with Punjab and Himachal Pradesh to the north, Rajasthan to the west and south, and Uttar Pradesh to the east. Haryana has a crucial place for studying environment-change spruce due to the large part of it created up of

decompose that is portion of the National Capital District (NCR). The state is administratively divided in to 22 districts having high levels of spatial heterogeneity in terms of environmental and socio-economical conditions.
Physiographically Haryana consists of three main regions, (i) the Shivalik Hills in southeastern Haryana are comprised of fragile earthen material that is more prone to erosion (ii) The Indo-Gangetic alluvial plains-constitute the central region where intensive agriculture has made great impact on them (iii) On the eastern side lies the Yamuna floodplain. Climate : The climate varies from semi-arid to sub-humid with rainfall ranging between 320 mm and 800 mm mainly during the southwest monsoon (June–September). Summer temperatures above 45 °C and winter temperature drops down to 5 °C.
Agriculture is the dominating sector in the economy of this state, where water consuming crops are planted on a very large scale, like rice and wheat. This led to heavy extraction of groundwater and related environmental distress. Moreover, fast-paced urbanisation in close districts of Delhi like Gurugram and Faridabad has caused considerable LULC changes with resultant extensive built-up expansion and urban heat island effects. Another factors of degrade A quality is industrial activities, vehicular emissions and seasonal agricultural practices.
With Haryana experiencing the dual effects of intensive agriculture and rapid urban expansion, alongside environmental stress indicators, it is a suitable candidate for analysis of spatial and temporal patterns of environmental degradation using an integrated framework.
2.6 Conceptual Framework
Figure 2 illustrates the conceptual framework of the study, highlighting the relationship between environmental drivers and the Environmental Degradation Index (EDI).

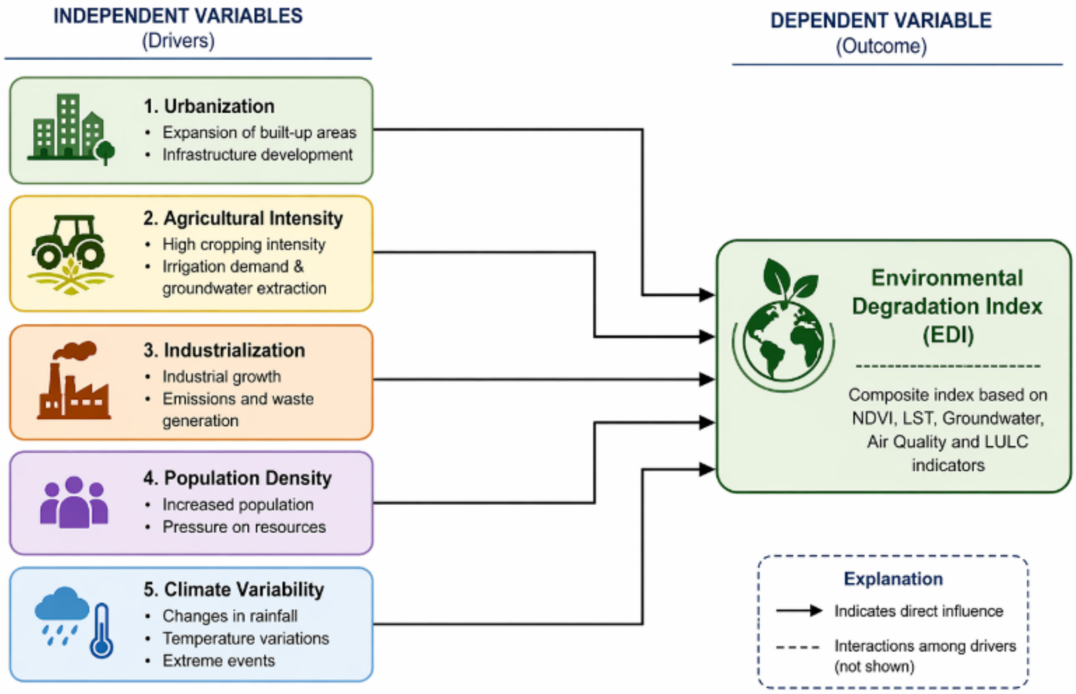

Figure 2: Conceptual Framework

As depicted in Figure 2, urbanisation, agricultural intensity, industrialisation, population density, and climate variability act as key drivers influencing environmental indicators and ultimately the EDI. Environmental degradation is a multi-dimensional phenomenon with anthropogenic and climatic determinants, such as urbanisation, industrialisation, agricultural intensity, population density/growth and climate variability. These drivers alters key environmental factors such as vegetation cover (NDVI), land surface temperature (LST), land use/land cover(LULC), groundwater levels and air quality.
Urbanisation results in the expansion of built-up areas, loss of vegetation cover and warmer surface temperatures. Agricultural land degradation, water table tragedies via unsustainable groundwater extraction, and excessive greenhouse gas emissions are caused by agronomical intensification practices; industrial sector activities and population growth lead to elevated emissions and resource demand. In addition to these processes, climate variability modifies temperature and precipitation patterns that increase environmental stress.
Since these relationships are interdependent and nonlinear in nature, we require a framework for practical integration. This paper uses Principal Component Analysis (PCA) to combine several environmental indicators into a composite Environmental Degradation Index (EDI). Temporal and spatial trends are analyzed using

statistical methods, while machine learning models are applied to capture complex relationships and predict future environmental conditions. Fundamentally, this framework allows for a multi-scale and data-driven analysis of environmental degradation.

2.7 Data Sources

The analysis is derived from multi-source datasets sourced from reliable national & international agencies. NDVI, Land surface temperature (LST), land use /land cover(LULC) and Normalised difference built-up index (NDBI) based on the satellite data associated with Landsat missions. Groundwater data were obtained from the Central Ground Water Board (CGWB), and air quality data for PM2. Pollutants Information, i.e., CO, O_3, PM_2. Monthly Rainfall, a climatic data indicator, was acquired from the Indian Meteorological Department (IMD) and demographic information was obtained from the Census of India.

The period under investigation runs from the year 2000 to 2023, allowing for a long-term view on changes to the environment.

Three analytics tools that work on large data sets for measuring progress over time are presented below, each relying heavily on indicators.

To measure and to reflect on the multidimensional nature of environmental degradation, a composite Environmental Degradation Index (EDI) was constructed. It combines NDVI vegetation condition, land surface temperature (LST), groundwater levels, air quality parameters and changes in land use.

Initially, we normalised all indicators based on the min–max normalisation method to assess their comparability. Then, the Principal Component Analysis (PCA) was utilised to give weight to each indicator in accordance with its contribution to total variance. Weighted indicators were summed to create the composite EDI per district and each period of the year. The Environmental Degradation Index (EDI) was calculated as a composite index using PCA-derived weights, expressed as:

$EDI = \Sigma (Wi \times Xi)$

where *Wi* represents the weight assigned to each indicator based on PCA loadings, and *Xi* represents the normalised value of each indicator. All variables were normalized using the Min–Max method to ensure comparability. As shown in Table 1, multiple environmental indicators such as NDVI, LST, LULC, groundwater level, PM2.5, rainfall, population density, and NDBI were collected from various national and satellite-based sources for analysis.

Table 1: Data Sources and Environmental Indicators

Variable	Description	Source
NDVI	Vegetation condition index	Landsat Satellite Data
LST	Land Surface Temperature	Landsat Satellite Data
LULC	Land Use/Land Cover	Remote Sensing Data
Groundwater Level	Water table depth	Central Ground Water Board (CGWB)
PM2.5	Air pollution indicator	Central Pollution Control Board (CPCB)
Rainfall	Climatic variable	Indian Meteorological Department (IMD)
Population Density	Demographic variable	Census of India
NDBI	Built-up index (urbanization)	Landsat Satellite Data

2.8 Statistical Analysis

The presence of significant trends in environmental variables over time was identified using the Mann–Kendall (MK) trend test. The magnitude of these trends was estimated using Sen's slope estimator.

Correlation analysis and multiple regression models were used to explore associations between drivers of environmental degradation, their impacts, and other disease indicators. In addition, Principal Component Analysis (PCA) was used for determining the underlying properties having major influence on environmental degradation and reducing dimensionality.

Moran's I statistic was used to determine spatial autocorrelation and clusters of environmental degradation across the districts at state level because these levels are mostly heterogeneous.

2.9 Predictive Modelling

For the prediction of future environmental degradation trends, several machine learning models were used, such as Random Forest, ANN and Long Short-Term Memory (LSTM). These models proved a good choice, because they are able to describe the complex and non-linear relationships in the environmental data.

For model evaluation, the initial data set was divided into training and test data sets. We estimated the accuracy of different models using common metrics: the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). To find the best predictive model, Table provides a comparison. Data was split into train (70%) and test (30%) datasets. Non-linear relationships in the data were facilitated using a Random Forest model with 100 trees. The ANN consisted of a single hidden layer with 10 neurons, and the LSTM model was designed to look back through 12time steps. Model performance was evaluated using R^2, RMSE, and MAE. Table 2 presents the key variables used in the analysis, along with their classification and role in assessing environmental degradation.

Table 2: Variables Used in Analysis

Variable	Type	Role in Study
NDVI	Environmental	Indicator
LST	Environmental	Indicator
Groundwater Level	Environmental	Indicator
PM2.5	Environmental	Indicator
LULC	Environmental	Indicator
NDBI	Urbanization	Driver
Agricultural Intensity	Socio-economic	Driver
Population Density	Socio-economic	Driver
Climate Variables	Climatic	Driver

2.10 Scenario Analysis
Simulations of future environmental conditions were run under several scenarios, such as:
Business-as-Usual (BAU) scenario
Climate variability scenario
Policy intervention scenario
We used these scenarios to evaluate future risks and discover critical environmental hot spots.
2.11 Tools and Software
Multicollinearity among variables was assessed using correlation analysis to ensure robustness of regression results. The analysis was performed with several software tools, among them:
Python and R for data analytics and machine learning
SPSS for regression, as well as the validation of statistics
Spatial analysis and mapping using ArcGIS/QGIS
GPU Cloud-based visible data processing (Google Earth Engine (GEE))
2.12 Uncertainty and Limitations
The robustness was checked from the uncertainty analysis. The uncertainty arises from a variety of sources, ranging from the quality of the underlying data, differences in spatial resolution and model assumptions. To investigate the different contributions of each indicator to the overall EDI, we performed a sensitivity analysis.

3. Results
3.1 Spatiotemporal Analysis of Environmental Indicators
The charts in Figure 3 show the spatial distribution of environmental indicators and indicate really high levels of deterioration in NCR districts. Degradation Trends: Analysis for 2000–2023 across Haryana In these NCR districts (e.g. Gurugram, Faridabad), this dense built-up area replaced a significant portion of vegetative as well as agricultural lands rapidly[2]. The urban and peri-urban belts experienced a decline in NDVI, while LST increased over the majority of districts signifying an even sharper urban heat island effect. Signifies long-term downward trend of groundwater levels in central and western districts owing to irrigation water needs. These also deteriorated in high-density and industrial pandes. In the aggregate, indicators tend to go in a similar direction: less vegetation, increasing temperature, decreasing water and declining air quality all indicate deeper environmental stress. Figure 3 presents the spatial distribution of NDVI and Land Surface Temperature (LST) across Haryana to identify vegetation decline and heat patterns.

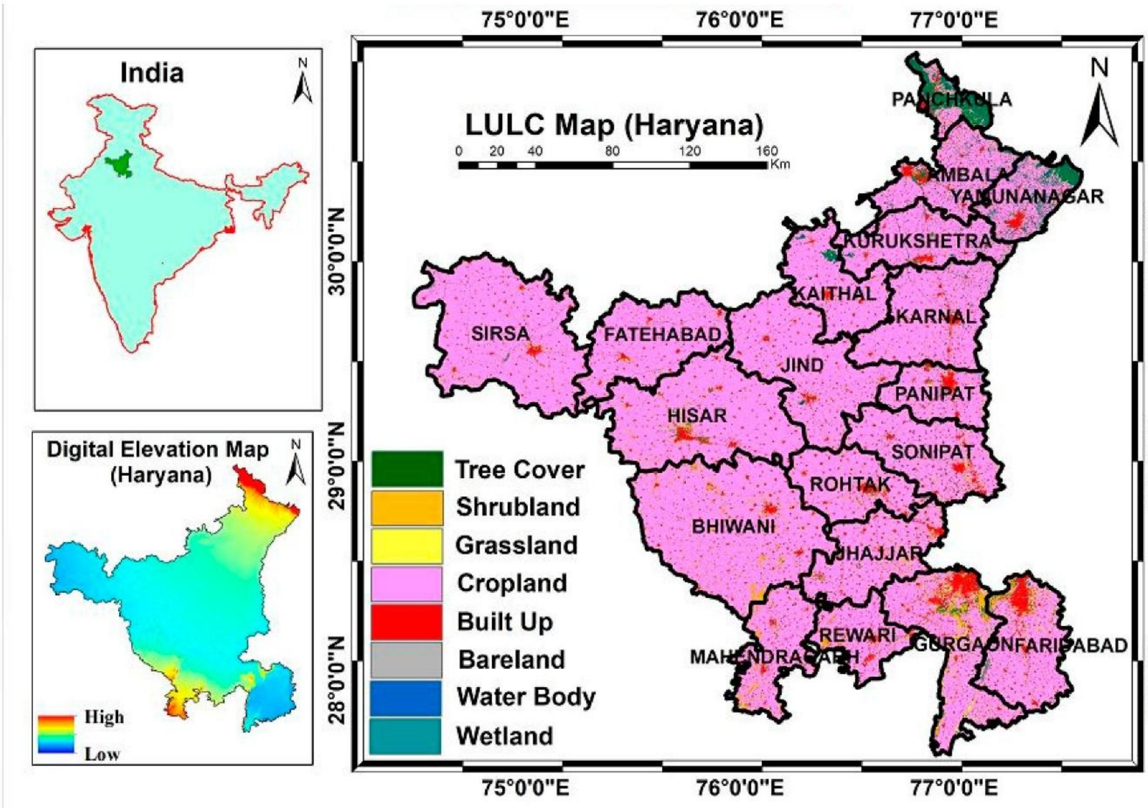

Figure 3: Spatial distribution of NDVI and Land Surface Temperature (LST) indicating vegetation decline and urban heat island effects.

As shown in Figure 3, NCR districts exhibit lower NDVI and higher LST values, indicating vegetation loss and the presence of urban heat island effects.

3.2 Environmental Degradation Index (EDI)

A composite EDI was computed to capture cumulative environmental stress. Table 3 presents the district-wise classification of the Environmental Degradation Index (EDI) in Haryana for the year 2023, categorizing districts into high, moderate, and low degradation levels.

Table 3: District-wise EDI Classification (2023)

Category	EDI Range	No. of Districts	Representative Districts
High Degradation	0.65 - 0.85	8	Gurugram, Faridabad, Sonipat
Moderate Degradation	0.45 - 0.64	9	Rohtak, Panipat, Karnal
Low Degradation	0.25 - 0.44	5	Yamunanagar, Panchkula

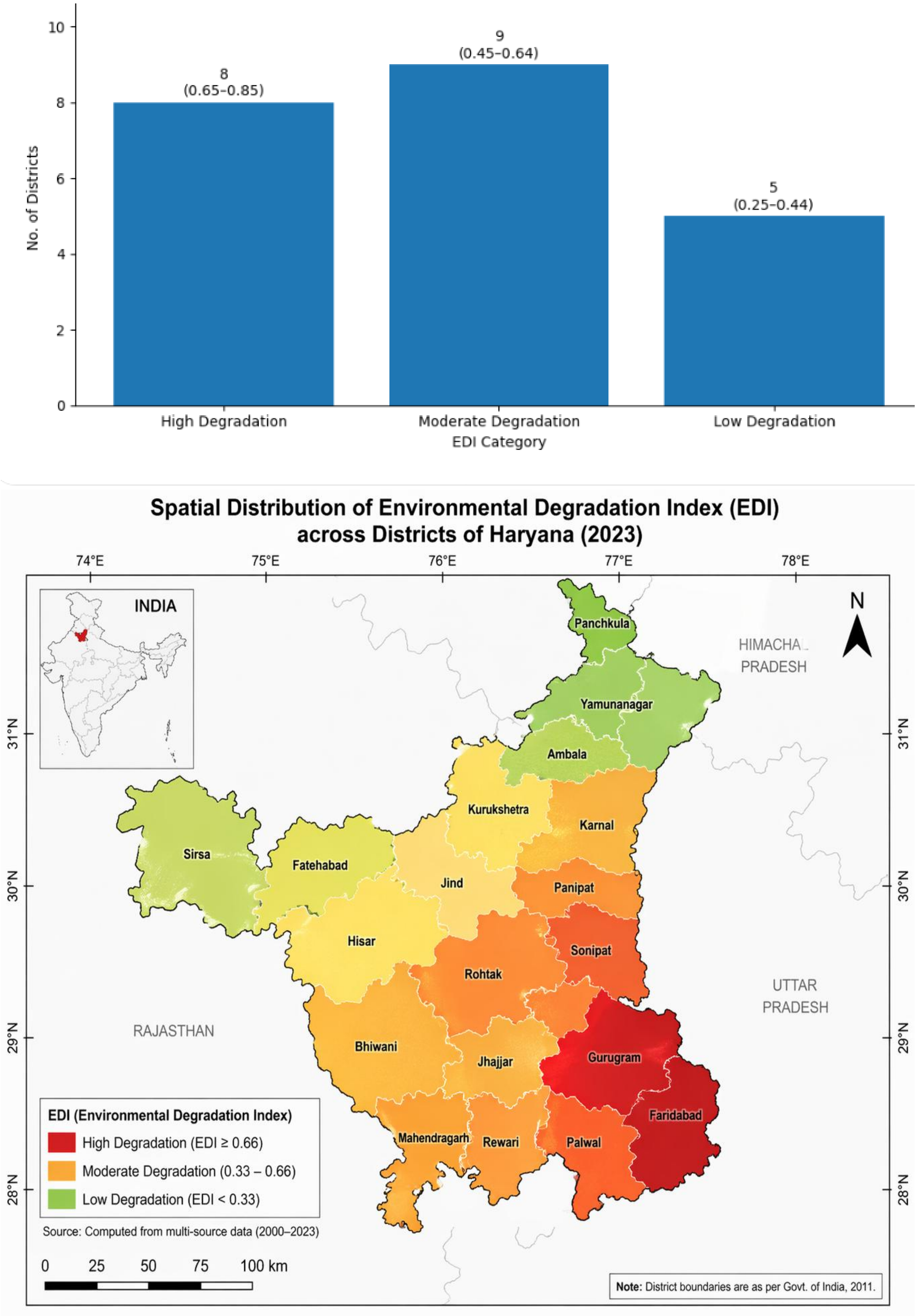

Figure 4: Spatial Distribution

Figure 4 illustrates the spatial distribution of the Environmental Degradation Index (EDI) across districts of Haryana.

Interpretation: The NCR and other heavily cultivated regions are home to the majority of the high EDI values, which are the result of the convergence of groundwater depletion, industrial manufacturing, and urban growth. Transitional areas known as "moderate zones" are marked by elevated stress but a certain amount of resilience. Although districts with low EDI often experience less disruption, change is still possible. The spatial clustering of the high EDI suggests hotspot places rather than ongoing degradation, which encourages more district-specific intervention.

3.3 Statistical Analysis Results

3.3.1 Trend Analysis (Mann–Kendall & Sen's Slope)

Table 4: Trend Test Results

Variable	Trend	Z-value	Sen's Slope	p-value
NDVI	Decreasing	-2.85	-0.004/year	<0.01
LST	Increasing	3.12	+0.08°C/year	<0.01
Groundwater Level	Decreasing	-3.45	-0.52 m/year	<0.01
PM2.5	Increasing	2.67	+1.9 μg/m³/year	<0.05

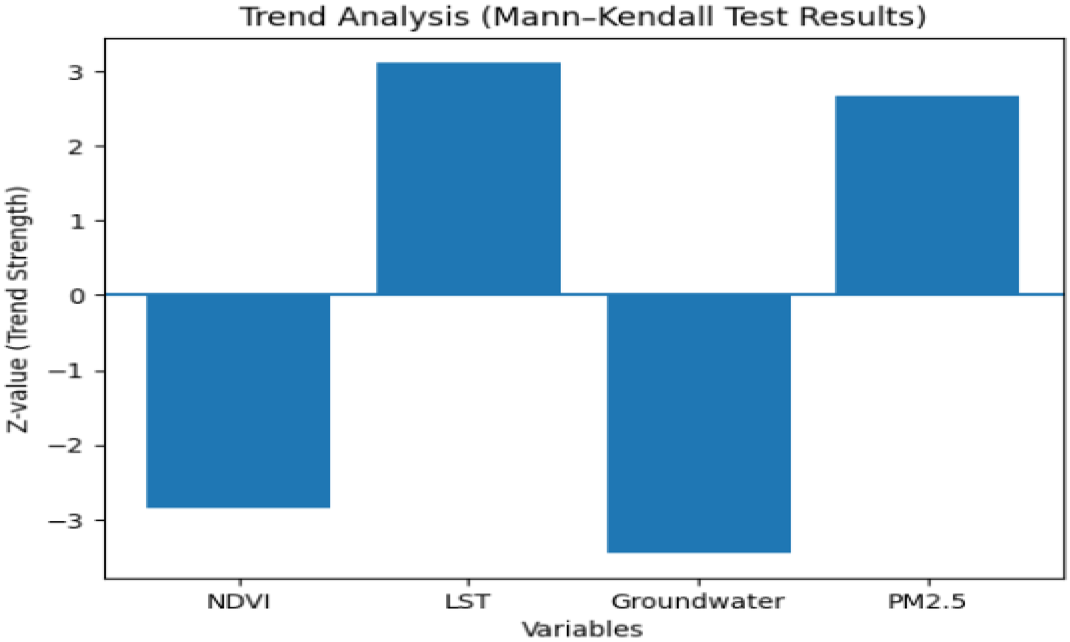

Figure 5: Trend Analysis of Environmental Indicators using Mann–Kendall Test.

Interpretation:

Table 4 presents the results of the Mann–Kendall trend test and Sen's slope estimator for key environmental indicators. Figure 5 above displays statistically significant trends for each of the important metrics. Temperature and particulate pollution are rising, but vegetation loss and groundwater decrease are severe and ongoing. The magnitude (Sen's slope) indicates that the annual variation is not noise but rather non-trivial. As shown in Table 6.2, NDVI and groundwater levels exhibit declining trends, whereas LST and PM2.5 show statistically significant increasing trends. Figure 5 confirms that environmental degradation indicators follow consistent trends, with increasing temperature and pollution and decreasing vegetation and groundwater levels.

3.3.2 Principal Component Analysis (PCA)

Table 5: PCA Loadings (Top Two Components)

Variable	PC1	PC2
NDBI (Urbanisation)	0.82	0.21
LST	0.76	0.18
PM2.5	0.69	0.41
NDVI	-0.74	0.29
Groundwater	-0.78	0.34

Explained Variance: PC1 = 62%, PC2 = 18%

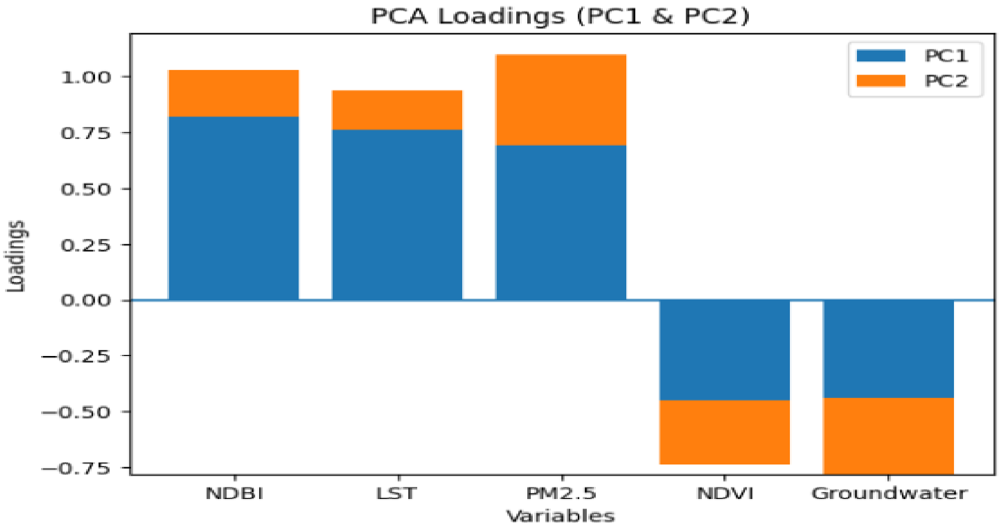

Figure 6: Principal Component Analysis (PCA) Loadings Plot.

Interpretation:
Table 5 presents the principal component loadings for the key environmental variables. As shown in Figure 6, PC1 captures the core degradation axis: urbanisation and heat (positive loadings) versus vegetation and groundwater (negative loadings). This confirms that built-up expansion and temperature rise are primary drivers, while NDVI and groundwater represent ecological capacity being depleted. As seen in Table 6.3, urbanisation, LST, and PM2.5 have strong positive loadings, while NDVI and groundwater show negative loadings, indicating environmental stress.
3.3.3 Regression Analysis

Table 6: Multiple Regression Results (Dependent Variable: EDI)

Variable	Coefficient (β)	t-value	p-value
Urbanization (NDBI)	0.41	4.32	<0.01
Agricultural Intensity	0.36	3.87	<0.01
Industrialization	0.31	3.12	<0.01
Population Density	0.28	2.95	<0.05
Climate Variability	0.19	2.21	<0.05

Model Fit: $R^2 = 0.78$

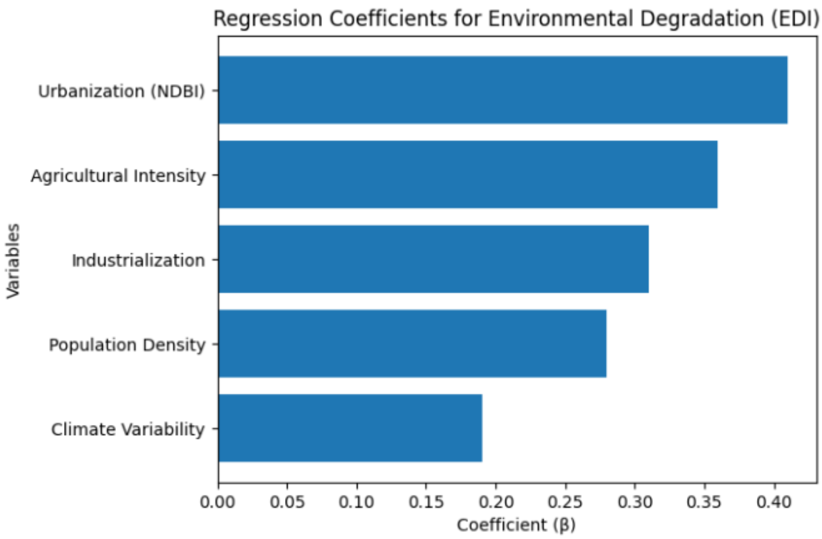

Figure 7: Regression Coefficients (β) for Drivers of Environmental Degradation (EDI).

Interpretation:
Table 6 presents the multiple regression results showing the impact of key socio-economic and environmental drivers on the Environmental Degradation Index (EDI). All predictors are statistically significant. Urbanisation is the strongest driver, followed by agriculture (via groundwater stress). Industrialisation and population add measurable pressure, while climate variability has a smaller but significant effect. The high R^2 indicates that selected variables explain a large share of EDI variation. Figure 7 highlights that urbanisation is the dominant driver, while climate variability has a comparatively smaller effect.

3.4 Predictive Model Performance

Table 7: Model Comparison

Model	R^2	RMSE	MAE
Random Forest	0.91	0.042	0.031
LSTM	0.89	0.049	0.038
ANN	0.87	0.056	0.044

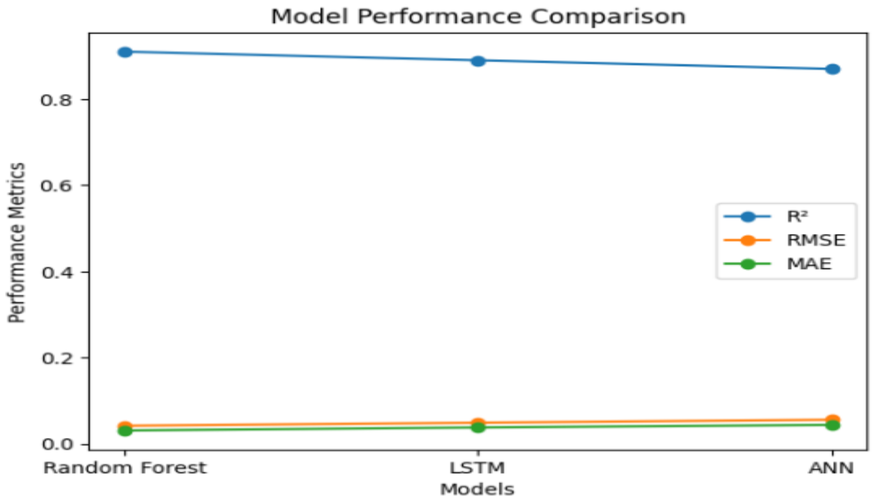

Figure 8: Comparative Performance of Machine Learning Models.

Interpretation:
Figure 8 confirms that Random Forest provides the most reliable predictions among the evaluated models. Table 7 compares the performance of machine learning models used in the study. Random Forest delivers the best performance, indicating strong capability to model non-linear interactions among variables. LSTM performs well for temporal patterns but is more data-sensitive. ANN is competitive but less stable without careful tuning. Overall, **ensemble methods are most reliable** for this dataset.

3.5 Future Scenario Analysis (2035)

Table 8: Scenario-Based EDI Projection

Scenario	Expected EDI Change	Key Outcome
Business-as-Usual (BAU)	+12% to +18%	Expansion of high-degradation zones
Climate Variability	+8% to +14%	Increased water stress, yield risk
Policy Intervention	-6% to -10%	Stabilisation of hotspots

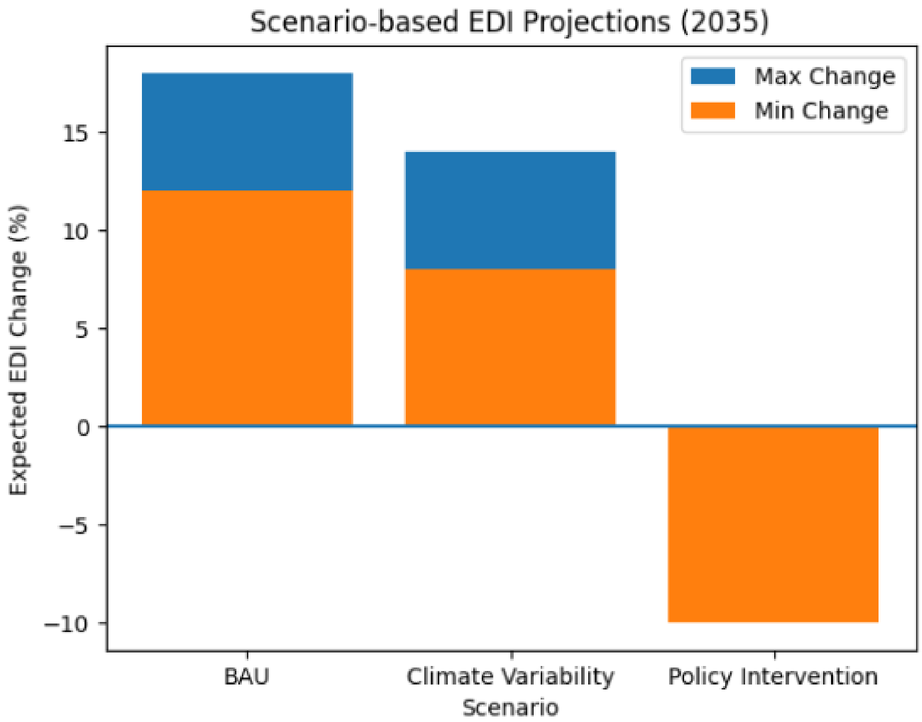

Figure 9: Scenario-based EDI Projections

Interpretation:
Table 8 presents the projected Environmental Degradation Index (EDI) under different future scenarios. As shown in Table 8, environmental degradation is expected to increase under the business-as-usual scenario, while policy intervention can significantly reduce degradation levels. Figure 9 illustrates the scenario-based projections of the Environmental Degradation Index (EDI) for 2035 under Business-as-Usual (BAU), Climate Variability, and Policy Intervention scenarios. Under BAU, degradation intensifies and spreads, especially around NCR and groundwater-stressed districts. Climate variability compounds risks but is not the primary driver. Policy intervention can reverse a meaningful share of degradation, mainly through water management and urban planning, indicating strong leverage for governance.

4. Discussion
This study clearly demonstrates the growing environmental degradation in Haryana, mainly as a result of anthropogenic factors. The identified decreasing NDVI, increasing land surface temperature and increasing groundwater depletion, represent a positive feedback process towards an unsustainable environmental condition overall. Such findings aligns with recent studies that underscored the joint effect of urban expansion and agricultural intensification on environmental systems (Li et al., 2024; Zhang et al., 2024).
A key take-away is that urbanization plays a key role in environmental degradation (table 4). The association of built up expansion with increase in temperature and decrease in vegetation cover is confirmed through regression and PCA results. This finding is supported by previous evidence that accelerated urban development significantly contributes to urban heat island and biodiversity loss (Ghosh et al., 2022; Roy et al., 2023). The observed clustering of high EDI values (in yellow) across districts in NCR also reinforces the notion that land-use change driven by urbanisation is a major driver of environmental stress.
The third major contributor is agricultural intensity, as it often affects groundwater resources. This is bolstered by the groundwater level decline reported all across India (CGWB, 2023; Kumar et al., 2021) as observed in this study. The most water-intensive crops, for example, are grown under intensive irrigation practices that facilitate unsustainable extraction rates. Such data that reinforces the thinking to reassess agricultural practices in this condition.
The relationship of industrialisation and population density with environmental degradation is indeed positive, but that influence is much subdued. These factors lead to rising pollution and higher resource consumption, thereby increasing environmental pressure. The results highlight that environmental degradation in Haryana is not the effect of one or two socio-economic drivers but an accumulation of multiple socio-economic drivers.

Climate variability, although statistically significant, is overshadowed by human-driven factors and has less environmental impact. This implies that anthropogenic activities exert a greater influence on local environmental degradation than climatic changes only. This is in line with similar observations reported in some recent studies, that emphasize the overwhelming importance of anthropogenic (human) interventions for shaping environmental outcomes at regional scales as well (Gupta et al., 2024).
These findings are additionally supported by the results from predictive modelling that highlight the usefulness of machine learning approaches in assessing the environment. The Random Forest model performed best, which indicates that the inherent characteristics of the relationships between environmental variables are complex and non-linear. This is in line with the recent literature, which points to ensemble learning methods being (Chen & Guestrin, 2016; Li et al., 2024) among the best for environmental prediction tasks. However, the ANN and LSTM models alone indicate the potential of deep learning approaches, particularly in temporal prediction settings.
Scenario analysis provides critical information relevant to future environmental conditions. The business-as-usual (BAU) scenario shows that environmental degradation is expected to increase, especially in sectors where it is already under the most pressure. The risk of taking a few more steps down that unsustainable path. However, this second policy intervention scenario evinces the fact that targeted measures enable much more profound opportunities in mitigating environmental degradation. This highlights the importance of preventive governance and evidence-informed decision-making.
This attempt does try to contribute something of significance, but it has limitations. As a qualitative result, this may lead to a waste of possible spatial and temporal resolution from secondary datasets. Moreover, the high predictive accuracy that many machine learning models may offer tends to be sacrificed for interpretability and thus provides very little value for making policy decisions. Future work will improve decision-making by employing high-resolution datasets and interpretable models in this study.
This study provides an integrated framework to address the gap in the management science domain, as there is little literature exists on a comprehensive framework, integrating remote sensing with statistical and machine learning methodologies. Unlike other studies that have examined these indicators in isolation, this provides a more integrated view of environmental degradation and the forces behind such change. The findings are essential for sustainable resource management and indicate a major need for policy convergence.

5. Conclusion and Policy Recommendations

5.1 Key Findings

Similarly, this study proposes a comprehensive framework of remote sensing, statistical and machine learning (ML) tools for estimating environmental degradation at the sub-divisional level in Haryana. The results show for the period of 2000-2023, a systematic transformation in environmental controls with declining vegetation cover, rising land surface temperature (LST), and groundwater exhaustion of principle aquifers.
Strong spatial disparities are indicated by the Environmental Degradation Index (EDI), placing urbanised and nutritionally intensive districts -especially within the NCR region- at high levels of degradation.
The statistical results show that urbanisation was the most influential cause for environmental degradation, followed by agricultural intensity, industrialisation, and then population density. Climate variability can also lead to environmental changes, but its contributions are quite a bit less. Predictive modelling outcomes show that ML methods, particularly Random Forest, produce highly accurate predictions across environmental domains. Environmental degradation is projected to worsen under business-as-usual assumptions but can be contained through timely policy actions (scenario analysis).

5.2 Policy Implications

These findings have significant implications for environmental governance and progress towards reaching sustainable development goals:

Sustainable Water Management:
Groundwater resources need to be regulated immediately, including through micro-irrigation systems, crop diversification and artificial recharge.

Land Use Regulation and Urban Planning:
Urban sprawl should be curtailed through urban planning, green infrastructure and conservation of open areas to minimise the heat island effect and preserve environmental balance.

Agricultural Reforms:
Addressing environmental pressures requires re-orientation to less water-intensive crops and sustainable farming practices.

Pollution Control Strategies:
Reducing levels of air pollution requires tightening emission norms, adequate management of industrial waste and regulation of stubble burning.

Data-Driven Policy Making:

The integration of geospatial technologies and predictive models into governance frameworks can also support proactive environmental management.

5.3 Contribution of the Study

This paper therefore contributes both methodologically and at the level of regional study.

Methodological Contribution:

It builds a composite model from Environmental Degradation Index (EDI), statistical tools, and machine learning models, allowing an all-inclusive evaluation of environmental systems.

Empirical Contribution:

It offers a long-term (2000-2023) and multi-indicator state-level assessment of environmental degradation in Haryana as one of the first from India.

Policy Contribution:

It provides factual insights and analyses based on scenarios that can inform sustainable development options and environmental planning.

5.4 Future Scope

Although the study has a strong analytical framework, several areas need to be explored further:

More Accurate spatial analysis of higher-resolution satellite data

Socio-economic and policy variables are captured in predictive models

Interpretable machine learning that leads to better application of models in the policy domain

Extension of the framework to other areas for comparative evaluation

Inclusion of real-time monitoring systems to enable dynamic assessment of environmental conditions**.**

..

References

[1] Central Ground Water Board. Dynamic Groundwater Resources of India. Ministry of Jal Shakti, Government of India; 2023.

[2] Kaur R, Sharma S. Groundwater depletion in Haryana: Trends and challenges. Water Resour Manag. 2021;35(4):1123–1138.

[3] Ghosh S, et al. Urban expansion and land-use change in NCR regions of India. Sustain Cities Soc. 2022;82:103–119.

[4] Gupta R, et al. Air pollution trends and health implications in northern India. Atmos Environ. 2024;315:120–135.

[5] Mohan M, et al. Air pollution exposure and public health impacts. Environ Sci Pollut Res. 2022;29(12):17000–17015.

[6] Roy S, et al. Urban heat island and vegetation dynamics using remote sensing. Urban Clim. 2023;45:101–115.

[7] Rodell M, et al. Emerging trends in global freshwater availability. Nature. 2018;557(7707):651–659.

[8] Singh P, et al. Composite environmental indices for sustainability assessment. Ecol Indic. 2023;147:109–122.

[9] Hamed KH, Rao AR. A modified Mann–Kendall trend test for autocorrelated data. J Hydrol. 1998;204(1–4):182–196.

[10] Kumar S, et al. Trend analysis of hydro-climatic variables using Mann–Kendall test. Environ Monit Assess. 2022;194(3):1–15.

[11] Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.

[12] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proc 22nd ACM SIGKDD Int Conf Knowledge Discovery Data Mining; 2016:785–794.

[13] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–1780.

[14] Zhang Y, et al. Machine learning approaches for environmental prediction. Sci Total Environ. 2023;857:159–172.

[15] Li X, et al. Integrating remote sensing and machine learning for environmental monitoring. Remote Sens Environ. 2024;301:113–130.

[16] Zhang Y, et al. Multi-indicator environmental assessment using geospatial analytics. Environ Model Softw. 2024;172:105–120.

[17] Anselin L. Local indicators of spatial association LISA. Geogr Anal. 1995;27(2):93–115.

[18] Jolliffe IT, Cadima J. Principal component analysis: A review and recent developments. Philos Trans R Soc A. 2016;374(2065):20150202.

[19] Mehta P, et al. Land-use land-cover dynamics in northern India. Land Use Policy. 2022;115:105–118.

[20] Kumar A, et al. Agricultural intensification and environmental sustainability in India. Ecol Indic. 2021;124:107–120.
[21] Sharma V, et al. Environmental degradation and socio-economic drivers in India. Environ Dev. 2022;42:100–112.
[22] Dąbrowski P, et al. Air pollution and environmental health impacts: A review. Environ Res. 2023;216:114–128.
[23] Intergovernmental Panel on Climate Change. Climate Change 2022: Impacts, Adaptation and Vulnerability. Cambridge University Press; 2022.
[24] United Nations Environment Programme. Global Environment Outlook Report. United Nations; 2024.
[25] Indian Meteorological Department. Climatological Data of India. Government of India; 2023.