Vol. 06, No. SI♣, pp. 35 –50 (2025)
ISSN: 2708-0757

JOURNAL OF APPLIED SCIENCE AND TECHNOLOGY TRENDS

www.jastt.org

Big Data-Based Remote-Sensed Intelligent Visual Analytics for Environmental and Earth Monitoring Systems

Pramoda Patro[1], M. Sowmya[2*], N. Srinivas[3], Subhranginee Das[4], Arul Kumar V P[5], R. Dhanapal[6]

[1]*School of Computer Science and Artificial Intelligence, SR University, Warangal, Telangana 506371, India, Pramoda.mtech09@gmail.com*
[2]*Department of Computer Science and Applications (MCA), SRM Institute of Science and Technology (FSH), Ramapuram Campus, Chennai, phdsowmya89@gmail.com*
[3]*Department of CSE, Vignana Bharathi Institute of Technology, Aushapur(V), Ghatkesar, Srinivas.bhaskar3@gmail.com*
[4]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, Telangana, India, subhranginee@gmail.com*
[5]*Computer Science and Engineering, Karpagam Institute of Technology, Coimbatore, Tamil Nadu, arulkumarssk@gmail.com*
[6]*Faculty of Engineering, Department of Computer Science and Engineering, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, dhanapalramasamy.p@gmail.com*

**Correspondence: phdsowmya89@gmail.com*

Abstract

The increasing accessibility of sensor-based earth observations, satellite imaging, and uncrewed aerial vehicle (UAV) data has made environmental monitoring a data-intensive field. However, the size, heterogeneity, and real-time nature of these datasets make it difficult to store, process, and interpret them effectively. This paper proposes a Big Data-driven approach to intelligent visual analytics for earth and environmental monitoring systems. With sophisticated machine learning algorithms, scalable data platforms, and remote sensing technologies, the framework enables automated feature extraction, anomaly detection, and spatiotemporal pattern recognition. Terabytes of satellite and sensor imagery can be analyzed more efficiently through the system's combination of Deep Learning-based Visual Interpretation and Distributed Computing Models (DL-VI-DCM), including Hadoop and Spark. Applications such as land cover classification, deforestation tracking, assessing the effects of climate change, disaster management, and water resource monitoring are all supported by the proposed methodology. To improve decision-making for scientists, environmental agencies, and politicians, a visual analytics dashboard is created to provide interactive maps, time-series trends, and predictive insights. Experimental findings from case studies on urban growth and flood-prone areas show that the framework can process high-resolution remote-sensed data in near real time with high accuracy and interpretability. This paper demonstrates the potential of integrating Big Data technology with intelligent visual analytics to provide flexible, scalable, and decision-focused solutions for earth system monitoring and global environmental sustainability. The proposed DL-VI-DCM achieves superior performance with 40–180 min of processing time, 92–95% classification accuracy, 89–94% anomaly precision, 55–110 GB/min scalability, 80–88% spatiotemporal accuracy, and 1.2–2.5 s response time.

Keywords: *Intelligent Visual Analytics, Big Data, Remote Sensing, Machine Learning, Earth Observation, Environmental Monitoring, Spatiotemporal Analysis.*

Received: October 18th, 2025 / Revised: November 28th, 2025 / Accepted: December 07th, 2025 / Online: December 11th, 2025

I. INTRODUCTION

Sensor-based Earth observation equipment that is easy to access has enabled continuous environmental monitoring as never before [1]. There are several places where one can always get spatiotemporal data. Some examples include UAVs, ground-based sensor networks, and high-resolution satellite imagery [2]. These figures could help us understand many intricate environmental problems, such as land-use changes, deforestation, urbanization, changes in water supplies, and climate change [3]. There are several challenges arising from the fact that this data is so large, varied, and often changing [4]. More complex computational frameworks are required to handle vast volumes of varied datasets and provide academics, policymakers, and environmentally conscious managers with relevant information when standard techniques for storing, processing, and analyzing data don't work [5].

The goal of this paper was to close the gap between the amount of data and the level of analytical skill, as dealing with continuous, high-resolution datasets, typical statistical procedures, and those that rely on human interpretation don't work well as data volumes grow [6]. It must employ integration

***SI♣*: Special issue - Remote Sensing based Intelligent Visual Analytics for Real-time Environmental and Earth Monitoring Systems**

methods to reveal functional patterns in the modern complex environmental data, including multispectral, hyperspectral, LiDAR, and radar data. Equally important is to convey the results in an understandable manner [7]. Decision-makers require unambiguous images that render analytics transparent to enable them to act promptly in domains like disaster management, climate change adaptation, and resource allocation [8]. Deep learning, machine learning, and distributed computing have made significant progress; yet, the systems they now have for monitoring the environment also have a long distance to travel [9]. Most systems do not have a wide variety of responses since they only function for certain tasks, like predicting the onset of floods, land cover classification, or deforestation assessment [10]. There are scalability issues, with few computers being able to handle terabytes of high-resolution images in real time [11]. Deep learning techniques can do predictions well but don't play nicely with visual analytics tools that allow users to easily do things with data [12]. It requires a platform that offers interactive visualization, scalable computation, and high-end analytics to overcome these issues [13]. This paper introduces a Big Data-enabled paradigm for smart visual analytics to help environmental and Earth-monitoring systems solve these issues [14].

The architecture rapidly processes enormous amounts of data on distributed computing platforms like Spark and Hadoop [15]. It also uses deep learning and machine learning to automatically identify features, problems, and trends in space and time [27]. An interactive visual dashboard is a giant leap in terms of making it simpler for people who need to know (including scientists, politicians, and environmental managers) to see the data, keep an eye on what's occurring, and make informed decisions [28]. It can apply the framework to map deforestation, track urban expansion, evaluate the influence that climate change has on things, react to emergencies, and categorize land cover [29]. The article speculates that Big Data-based visual analytics could revolutionize the way it is monitoring the environment because they are simple to use, scalable, and automated [31]. The system presents a solid, flexible, and decision-oriented solution to urgent global environmental issues by combining advanced computational analysis with pragmatic decision-making. This accelerates and enhances evaluations. The DL-VI-DCM framework exhibits technical superiority over traditional distributed remote sensing systems through its fusion of deep learning-driven feature extraction and high-efficiency distributed computing. It employs advanced convolutional and transformer-based architectures within a Hadoop–Spark ecosystem to enable scalable, parallel visual analytics across multi-source and high-resolution remote-sensed datasets. The framework leverages in-memory data processing, optimized task scheduling, and adaptive load balancing to achieve significant gains in computational throughput and latency reduction. Its architecture supports automated spatial–temporal feature learning, enhancing precision in environmental classification, anomaly detection, and terrain change analysis.

II. Related Works

Geographical data is essential for creating cities, monitoring the environment, traveling, and coping with disasters. It needs GIS to do this. More and more satellites, sensors, and mobile devices are generating large amounts of geographic data that need better ways to be analyzed. Data-Driven GIS Analysis (DD-GISA) is an approach that uses GIS, AI, statistics, and machine learning simultaneously [16]. This paper proposes it. DD-GISA helps businesses utilize their resources more effectively and solve challenging geographical problems that other methods can't. Finding patterns, trends, and connections in large datasets allows people across many fields to make better decisions and predictions.

Storing, interpreting, and making meaning of remote sensing data is a vast data problem, given the many sensors and satellites. This paper introduces High-Performance Remote Sensing Processing (HPRSP) as a method to improve the efficiency of processing heterogeneous data from several sources [17]. It utilizes HTC and HPC to achieve this. HPRSP overcomes generalizability concerns, timeliness, and security through the capability of researchers to analyze massive remote sensing datasets that can be scaled to any size. The faster outputs of the approach could assist scientific analysis with high computing needs, resource management, and environmental consideration.

Cloud computing remote sensing systems (CCRS) can effectively store, analyze, and scale big volumes of remote sensing data. This paper presents Cloud-Enhanced Remote Sensing Analysis (CERSA), a method that supports complete data processing, insight generation, and decision support. CERSA is able to address the increasing requirement for flexible, consistent remote sensing systems that may scale with users' needs [18]. CCRS infrastructure and advancing technologies such as AI, big data, and the Internet of Things make this possible. This approach is less difficult for schools and the government to implement, as it is more efficient in tracking the environment, city management, and natural resource study. Deep learning has revolutionized the way scientists create land use and cover maps on our planet.

Deep Learning Semantic Mapping (DLSM) is introduced in this paper, where semantic segmentation networks are used to simulate multi-scale effects, capture contextual relationships, and retrieve spatial and semantic features [19]. DLSM simplifies moving information around and handles location changes. Applications are farming, clustering trees and species, and detecting building boundaries. The method is better than previous approaches to mapping land cover and tracking resources since it uses DLSM and large amounts of remote sensing information. This enhances the system's predictions, extracts information automatically, and allows researchers to explore on a larger scale. GIScience and remote sensing are capable of educating students significantly about the environment and natural resources.

Environmental sensor-based monitoring is introduced in this paper as a core aspect of the Environmental GIScience & Remote Sensing Framework (EGIRSF), which integrates geo-analysis, geo-computation, and geo-modeling from GIScience. EGIRSF's capability of processing vast amounts of heterogeneous data rapidly might aid EIAs, natural resource management, and forecasting [20]. This approach to considering problems that entail extensive processing and analysis of data

renders environmental and resource science more precise, scalable, and useful. AI and big data have simplified using GIS to plan cities, track the environment, and manage resources more than ever before.

The present work proposes AI-Enhanced GIS Analytics (AIGA), a paradigm that leverages AI algorithms for modeling and predicting spatial trends from large, heterogeneous datasets [21]. AIGA enhances existing GIS techniques through the application of learning-based pattern recognition, prediction modeling, and feature extraction. The approach leverages traditional GIS with new technologies that need a lot of data. It supports resource allocation, enhances analysis accuracy, and accelerates decision-making in inaccessible areas. Researchers have to study a broad spectrum of environmental information in order to learn about climate change.

This is the Climate Big Data Analytics (CBDA) page, which is how to learn about weather trends, how they affect infrastructure, and how natural resources are evolving. It employs sensors, IoT, and sophisticated data mining methods. CBDA assists individuals who have to cope with climate change and policymakers by overcoming issues with data that differs, is difficult to manage, and not necessarily current [22]. CBDA employs vast volumes of data to enable individuals to make fact-based choices, forecast future climate results, and solve environmental and social problems. The WLRTMS employs the integration of hybrid sensors (accelerometers and GNSS), data infrastructure, and analysis algorithms to track groundwater, vibration, and displacement changes [23].

WLRTMS enables it to maintain a close watch on landslides, seeing when they begin, how quickly they travel, and where they might travel. It simplifies managing disasters and alert systems for individuals. The WLRTMS system utilizes field sensor networks and large-scale data analytics to enable strong, adaptable, and scalable approaches to monitoring distant regions or terrain that is susceptible to landslides. Deep learning revolutionized the acquisition of sea ice (SIE) dramatically.

This paper presents Deep Learning Sea Ice Extraction (DLSIE), a framework that integrates deep learning, machine learning, and traditional segmentation techniques to process SAR and optical information. DLSIE simplifies the mapping of ice, which aids climate analysis, navigation, and GIS [24]. This technology dispenses with the constraints of existing image processing methods to facilitate automatic feature identification and scanning of extensive regions of the polar areas. In the future, it would like DLSIE to work with other forms of ice and in different scenarios. Urban informatics builds data-driven cities by integrating big data and the IoT.

This assists in transportation, environmental monitoring, and policy-making. This article explains the Smart City Urban Informatics Framework (SCUIF), which allows the useful information to be extracted from city data [25]. Examples of London and Seattle show that SCUIF allows us to see how cities change as they change. SCUIF integrates AI and machine learning to create models that predict the future, assist in dealing with climate change, and improve infrastructure. This enables cities to expand in a way that is beneficial to the environment and also makes government function more efficiently. In below Table I, shows the comparison of existing methods.

III. Methodology

The suggested method intends to create a single system of environmental monitoring through the integration of smart visual analysis, big data analytics, and remote sensing methodologies. The key concept of this framework is to use high-performance machine learning models and distributed computing platforms to automate the processing of data and discovery of patterns [32]. Data are sourced from different places to create a single dataset that includes, without modification, uncrewed aerial vehicle (UAV) activities, sensor feeds, and satellite pictures. The framework is modular, scalable, and intuitive, so it is capable of enabling and analyzing very large spatiotemporal datasets and providing useful knowledge to the field [33]. The system has four essential levels: data collection, preprocessing, analytics and visualization. The levels enable localization of features, identification of problems, and actionable decisions across diverse environmental domains.

A. Stage 1 – Multisource Data Acquisition and Preprocessing Module

The first stage in creating the DL-VI-DCM is to systematically collect and organize data from multiple environmental sources, as shown in Fig. 1. This is achievable through a combination of ground-based IoT sensors, satellite imagery from optical, radar, and LIDAR sensors, and UAV-based aerial photogrammetry. Two examples of distributed storage systems that store raw data are Hadoop, HDFS, and Spark. The framework employs multimodal feature alignment modules based on attention-guided convolutional and transformer networks to harmonize variations in spatial resolution, spectral range, and acquisition geometry across heterogeneous data sources. Feature maps from satellite imagery and UAV captures are normalized and co-registered using geospatial referencing and learned affine transformations, ensuring consistent spatial alignment. Subsequently, a multi-scale fusion layer aggregates complementary information—such as high-frequency texture details from UAV data and broad contextual cues from satellite imagery—within the Spark-based distributed processing pipeline. This enables parallel feature integration and reduces redundancy through adaptive weighting schemes optimized by gradient-based feedback.

This makes it easier to handle large datasets that could grow even larger. A preprocessing layer cleans and standardizes data by removing noise, filling in missing values, and adding comments. It's better for distributed machine learning to place the data in the right order. This step ensures that the deep learning and analytics modules that follow later can access high-quality data in a consistent format. This will enable us to monitor and predict the weather with great accuracy.

TABLE I. COMPARISON OF EXISTING METHODS

S.No	Focus Area	Methods	Main Focus	Data Type	Key Applications	Limitations
16	GIS & Data-Driven Analytics	DD-GISA	Integrates AI/ML with GIS to identify patterns and optimize resources	Spatial data from satellites, sensors, and mobile devices	Urban planning, environmental monitoring, and disaster management	Requires large, high-quality datasets; may struggle with heterogeneous data integration; computationally intensive
17	Remote Sensing Big Data & HPC/HTC	HPRSP	Uses HPC/HTC to improve computation and storage efficiency	Multisource heterogeneous remote sensing data	Earth observation, environmental monitoring, scientific analysis	High infrastructure cost; scalability issues with huge datasets; requires specialized expertise
18	CCRS for Remote Sensing	CERSA	Leverages cloud computing for scalable processing and analysis	Large-scale remote sensing datasets	Urban management, resource analysis, and environmental monitoring	Data security and privacy concerns; dependency on cloud availability; latency issues for real-time applications
19	Deep Learning for Land Cover Mapping	DLSM	Semantic segmentation for spatial & semantic feature extraction	Remote sensing images	Agriculture, building extraction, tree segmentation, and land use mapping	Model transferability across regions may be limited; requires annotated datasets; high computational demand
20	GIScience & Remote Sensing for Environment	EGIRSF	Combines geo-modeling, geo-analysis, and geo-computation with environmental monitoring	GIS datasets, environmental sensors	Natural resource management, environmental impact assessment	A complex framework can be challenging to implement, may require expert knowledge, and be sensitive to data quality
21	AI & Big Data in GIS	AIGA	Integrates AI with GIS for improved modeling, prediction, and analysis	Large, heterogeneous GIS datasets	Urban planning, resource management, and environmental monitoring	Performance depends on data quality, risk of overfitting, and high computational cost
22	Climate Change & Big Data	CBDA	Analyzes climate and environmental data using IoT & advanced analytics	Climate, environmental, sensor data	Climate modeling, mitigation planning, predictive insights	Heterogeneous and noisy data; high storage and computation requirements; predictive uncertainty
23	Landslide Monitoring	WLRTMS	Real-time monitoring of landslide movement with hybrid sensors and big data	GNSS, accelerometer, groundwater, displacement, vibration data	Disaster management, early warning, slope stability analysis	Expensive deployment; sensor maintenance in harsh environments; limited coverage area
24	Sea Ice Extraction	DLSIE	Combines classical, ML, and deep learning for ice mapping	SAR, optical remote sensing data	Climate analysis, navigation, GIS production	Limited labeled datasets; performance varies by sensor type and environmental conditions; computationally intensive
25	Urban Informatics for Smart Cities	SCUIF	Integrates IoT, AI, and big data for data-driven urban management	IoT sensor data, social media, urban metrics	Thoughtful city planning, mobility, environmental monitoring, and climate adaptation	Data privacy concerns, uneven sensor coverage, and integration challenges across heterogeneous urban systems

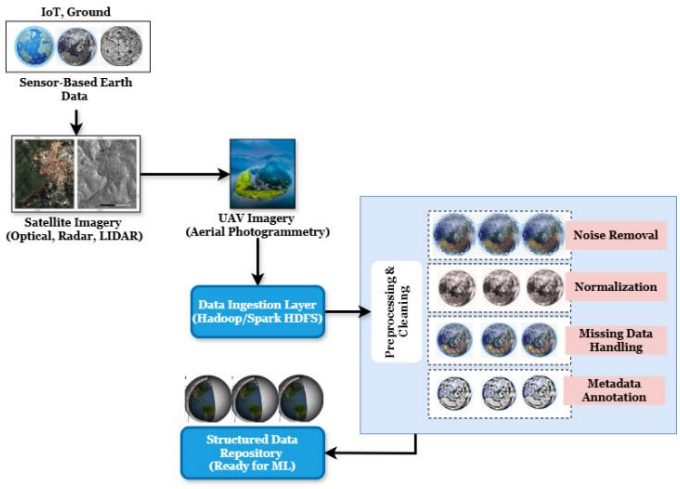

Fig. 1. Multisource Data Acquisition and Preprocessing Module.

This makes it easier to handle large datasets that could grow even larger. A preprocessing layer cleans and standardizes data by removing noise, filling in missing values, and adding comments. It's better for distributed machine learning to place the data in the right order. This step ensures that the deep learning and analytics modules that follow later can access high-quality data in a consistent format. This will enable us to monitor and predict the weather with great accuracy. Data fusion and preprocessing function $E_{gq}(u)$ is expressed in Eq. 1

$$E_{gq}(u) = \gamma_1 . \frac{(T_j . U_j)}{m} + \gamma_2 . \frac{(V_j . G_j)}{n} + \gamma_3 . \qquad (1)$$

This equation models the fusion of multi-source environmental data from sensors, UAVs, and remote sensing satellites.

T_j denotes data streams from IoT sensors, and U_j Their respective weight factors. V_j refers to UAV-collected imagery inputs, and G_j Denotes corresponding feature extraction parameters. S_l represents remotely sensed data from satellites,

and π_k their correction or calibration factors. The constants $\gamma_1, \gamma_2, \gamma_3$ denote scaling coefficients for each modality, while the total counts of the sensor m, UAV, $,n$ and satellite data sources q, respectively. A well-defined flow would begin with the data ingestion and preprocessing stage, where multisource satellite and UAV imagery undergo normalization, noise reduction, spectral calibration, and dimensionality reduction using PCA–ICA transformations within the Hadoop ecosystem. These preprocessed datasets are then seamlessly transferred into the distributed computation layer, where Spark manages parallelized deep learning operations—including convolutional feature extraction, attention-based fusion, and spatiotemporal inference—through in-memory task scheduling and adaptive resource allocation. The output from this computation layer feeds directly into the visualization and analytics module, which employs distributed rendering engines and GIS-based dashboards for real-time mapping, anomaly tracking, and environmental trend visualization. Distributed analytical computation layer $B_e(y, x, u)$ is expressed in Eq. 2

$$B_e(y,x,u) = \delta\ \vartheta_s \,.\, g_s\ \frac{\varepsilon^2 E_{gq}(y,x,t)}{\varepsilon y^2 + \varepsilon x^2} + \beta.\frac{\varepsilon E_{gq}(y,x,t)}{\varepsilon u} \quad (2)$$

This equation captures the distributed analytical process where spatial-temporal derivatives are computed over preprocessed data.

The spatial coordinates (y, x, u)at time u. E_{gq} is the fused dataset from the previous layer. g_s denotes the nonlinear analytical function applied by the sth distributed computation unit, while ε^2 represents its adaptive weight. The parameter ϑ adjusts the influence of temporal changes relative to spatial gradients. δ represents the global aggregation operator that consolidates node-level computations.

Deep learning prediction and visualization mapping $W_{pr}(y, x, i)$ is expressed in Eq. 3

$$W_{pr}(y,x,i) = \emptyset\ [\rho(V_m.B_{cm}(y,x,i) + c_m) \aleph\ L_m(y,z)] + \alpha_u.\nabla_{rq}\big(B_{cm}(y,x,i)\big) \quad (3)$$

This equation represents the deep learning-based predictive visualization process, where analytical outputs are passed through convolutional layers.

The spatial position $(y, z)$and time $i.B_{cm}(y, x, i)$refers to the analytical layer output, V_m and c_m are the weight matrix and bias term for the ith deep layer respectively, and ρ is the activation function. $L_m(y, z)$indicates the convolutional kernel applied to capture local spatial correlations, while α_u governs temporal sensitivity for predictive mapping. The symbol ∇_{rq} denotes the spatial gradient operator, $\aleph$ indicates convolution, and $\emptyset$ is the visualization transformation function. Each convolutional kernel functions as a localized filter that slides across the input feature map, learning spatial hierarchies such as edges, boundaries, vegetation textures, and terrain gradients from multispectral and UAV data. By stacking multiple convolutional layers with varying receptive fields, the framework progressively aggregates low-level structural cues into high-level semantic representations, enabling precise delineation of land-use classes and environmental anomalies. The integration of multi-scale and dilated convolution operations further strengthens the model's capacity to preserve spatial context while expanding the field of view, allowing accurate detection of small and large-scale environmental changes. Simplified sensor allocation function v^* is expressed in Eq. 4

This equation determines the simplest balance between gaining valuable sensor information and minimizing cost and bandwidth use.

$$v^* = argmax\ IH_j\ (v_j) - \delta\ D_j(v_j) - q\ c_j(v_j) - C_{\underset{b}{avai}} \quad (4)$$

This equation determines the simplest balance between gaining valuable sensor information and minimizing cost and bandwidth use.

v_j represents the allocation, term IH_j denotes the information gained from the sensor j, and $D_j(v_j)$is its cost. The coefficient δ scales the influence of cost, while q penalizes excess bandwidth usage when the total $c_j(v_j)$exceeds the available limit $C_{\underset{b}{avai}}$. Spark was chosen for large-scale remote sensing data processing in the DL-VI-DCM architecture owing to its in-memory processing, iterative computing performance, and scalability. Spark reduces disk I/O by keeping intermediate data in memory, speeding up deep learning-driven visual analytics and feature extraction jobs that need repetitive data access. Its Resilient Distributed Dataset (RDD) and DataFrame APIs enable real-time analysis of multispectral and UAV data streams using fault-tolerant, distributed data structures suited for batch and streaming operations. Spark's integration with deep learning frameworks like TensorFlowOnSpark and BigDL allows distributed model training and inference in the same computing environment. Data locality and resource usage across big clusters are optimized by its HDFS compatibility.

Pseudocode 1: Deep Learning-Based Visual Interpretation (DL-VI)

Input: Satellite image dataset X = {x_1, x_2, ..., x_n}, labels Y = {y_1, y_2, ..., y_n}
Output: Classified land cover map $\hat{Y}$

1. Initialize CNN parameters W^c, biases b^c, learning rate η
2. For each epoch e = 1 to E do
3. For each batch (x_i, y_i) in X do
4. Extract features f_i = Conv (W^c * x_i + b^c)
5. Apply attention weights α_i = Softmax(QK^T / $\sqrt{d_k}$)
6. Compute context vector c_i = Σ_j α_{ij} V_j
7. Concatenate multi-head outputs F = Concat(c_1, c_2, ..., c_h)
8. Predict output $\hat{y}_i$ = Softmax(W^oF + b^o)
9. Compute loss L = $\Sigma(y_i \log(\hat{y}_i))$
10. Update parameters $W^c \leftarrow W^c - \eta\nabla L$
11. End for
12. End for
13. Generate classified map $\hat{Y}$ = Argmax($\hat{y}_i$)
14. Return $\hat{Y}$
15. End

This pseudocode 1 outlines a Deep Learning-based Visual Interpretation model that processes satellite images using convolutional neural networks and attention mechanisms. It extracts spatial features, applies multi-head attention to focus on critical areas, and classifies land cover types efficiently through backpropagation, achieving high accuracy in environmental image interpretation. Domain specialists including environmental scientists, GIS analysts, and disaster response teams might review interface intuitiveness, interaction efficiency, and information clarity. Quantify user experience and decision-making effectiveness using usability measurements including task completion time, user accuracy rate, and System Usability Scale (SUS) ratings. Benchmark dashboard responsiveness under different data loads to demonstrate its capacity to handle large-scale, real-time geographic representations. By improving frame refresh rate, latency, and throughput under concurrent visualization applications, WebGL-accelerated rendering and distributed visualization caching merit integration.

B. Stage 2 – Intelligent Analytics and Distributed Processing Module (DL-VI-DCM Core)

The data collecting method requires many in-situ sensors, multi- and hyperspectral imagery, LiDAR, and UAV imagery. These datasets are hard to forecast from and poor [34]. They must be cleaned up first. Precision may be achieved by geometric correction, noise reduction, data standardization, and radiometric calibration. Fusion of data from several sources improves feature accuracy and reduce redundancy. When data is inaccurate or missing, statistical imputation and interpolation are used to correct errors [35]. The following modules for machine learning and analytics need datasets that have already been processed and can be easily fed into distributed computing platforms. This ensures it uses standardized, high-quality inputs for accurate spatiotemporal analysis and visualization.

Fig. 2, the primary portion of the analysis, shows how to manage massive data in an organized manner using distributed computing frameworks such as Spark or Hadoop MapReduce. Maybe can deal with enormous datasets quicker if you break them up into smaller pieces and process them all at once. It uses deep learning models including CNNs, RNNs, and autoencoders to detect features, model time, and detect anomalies. The spatiotemporal pattern recognition engine may extract useful information from raw data by detecting objects, segmenting land cover, and recognizing changes. The outputs include environmental maps organized by category, evaluations of unusual events, and models that forecast future events in this scenario, distributed computing and powerful visual analytics work together to identify real-world patterns and trends that are accurate and useful to many people. Then, the data is used to develop forecasts and monitor current events. Demonstrating scalability through multi-cluster experimentation on heterogeneous cloud and edge infrastructures—such as AWS EC2, Google Cloud Dataproc, or on-premise Spark clusters—would provide verifiable performance indicators across diverse computational environments. Metrics including linear speedup ratios, data throughput per node, latency under increasing workload, and energy consumption scalability curves can be systematically recorded to quantify system behavior as dataset size and node count increase. Additionally, evaluating cross-site synchronization efficiency and fault-tolerance performance under real network conditions would establish the robustness of distributed model coordination. Integrating these experimental validations with visual analytics of node utilization and task scheduling timelines would deliver concrete, data-backed evidence of the framework's scalability.

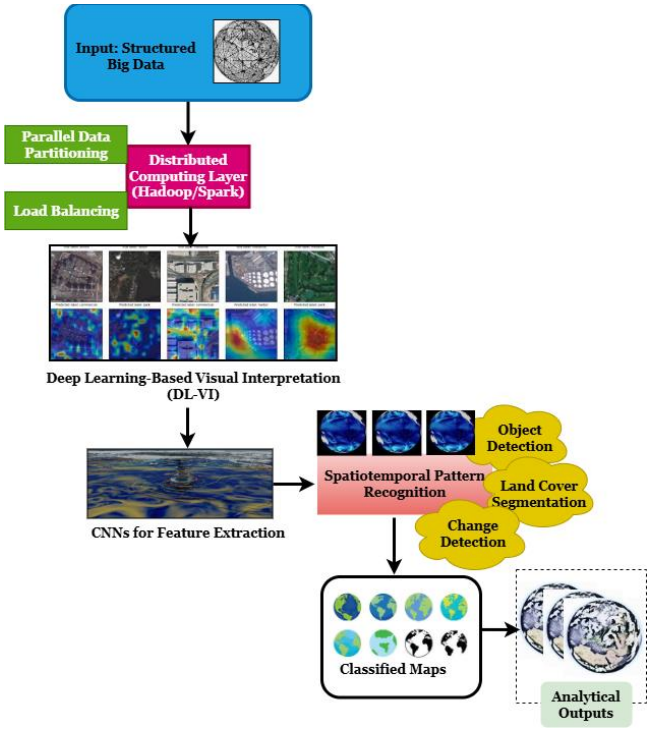

Fig. 2. Intelligent Analytics and Distributed Processing Module.

Ingestion rate & latency budget $A(u)$ is expressed in Eq. 5

$$A(u) = q_r(u) + \gamma_v(u) - \frac{1}{\Delta u}\log 1 + \theta_{log}(u) \qquad (5)$$

This expression quantifies the net data ingestion capacity by combining source throughput and per-UAV transfer bursts while penalizing accumulated latency.

In this, $q_r(u)$is the instantaneous data rate from sensor source r, $\gamma_v(u)$is the burst transfer rate contribution from UAV v, Δu is the measurement interval used for rate normalization, and $\theta_{log}(u)$is the accumulated network/system latency factor at time u. Missing-value imputation $Y(u)$ is expressed in Eq. 6

$$Y(u) = \partial L_{in}\, X_{ob}(u) + (1-\beta)\mu_\theta\, X_{ob}(u) \qquad (6)$$

This convex combination fills gaps by blending fast interpolation with a learned model prediction to handle complex patterns and context.

In this, ∂ is the blending coefficient between interpolation and model output, $L_{in}(\cdot)$ is a deterministic interpolation operator applied to observed entries $X_{ob}(u)$and $\mu_\theta(\cdot)$ is a trained parametric imputation model (parameterized by θ) applied to the same observed data. Noise suppression $Y(u)$ is expressed in Eq. 7

$$Y(u) = K(u)Y(u) + 1 - L(u)Y_{raw}(u)\,K(u) \qquad (7)$$

This update reduces measurement noise by blending the imputed/filtered estimate and the raw measurement using a time-varying gain.

In this, $K(u)$is the adaptive blending (Kalman) gain at time u, $Y(u)$is the imputed/estimated state (from the previous equation), $Y_{raw}(u)$is the newly received raw measurement, and $S(u)$is the measurement noise variance.

Indexing & storage cost vs. retrieval latency K is expressed in Eq. 8

$$K = \pi_e\, R(E) + \pi_j J(E) + \pi_p\, F_p\,(p;E) \qquad (8)$$

This cost function trades storage footprint and index size against expected query retrieval latency to choose storage and indexing strategies.

In this, scalar weights for storage π_e, index π_j, and query latency π_p respectively, $R(E)$Is the storage footprint, $J(E)$is the index size or the indexing overhead for E, and F_p is the expected retrieval time averaged $(p;E)$ over the query workload p when accessing the dataset E.

Pseudocode 2: Distributed Computing Model (DCM) using Spark + Hadoop

Input: Large-scale sensor data D = {d_1, d_2, ..., d_m}, cluster nodes N
Output: Anomaly score matrix A[m][t]

1. Initialize SparkContext(SC), Hadoop HDFS storage
2. Partition D into subsets D_n across N nodes
3. For each node n ∈ N do
4. Load D_n → RDD format
5. For each record $d_i \in D_n$ do
6. Compute feature vector v_i = MapReduce(d_i)
7. Calculate anomaly score $a_{it} = |\mu_t - d_{it}| / \sigma_t$
8. If $a_{it} > \theta$ then mark as anomaly
9. End for
10. Aggregate results $A_n = \Sigma(a_{it}) / |D_n|$
11. End for
12. Combine A = ∪A_n and normalize over time steps t
13. Generate spatiotemporal visualization V(A)
14. Return A, V(A)
15. End

This pseudocode 2 illustrates a Spark-Hadoop Distributed Computing Model for big environmental datasets. Real-time anomaly detection and decision-oriented environmental analysis are enabled by partitioning data among cluster nodes for parallel processing, statistical thresholds for anomaly scores, and spatiotemporal maps for aggregated findings.

C. Stage 3 – Visual Analytics and Decision Support Dashboard

Given the large dimensionality frequent inclusion of irrelevant or duplicate information in remote-sensed data, feature extraction and dimensionality reduction must be fast. PCA, ICA, and convolutional feature extraction reduce data while keeping spatial and spectral information. Feature selection improves classification, anomaly detection, and change detection. CNN-based deep learning extraction may create hierarchical visual representations [36]. These techniques simplify the visual analytics dashboard by combining processing power and data security. Scalable analysis for environmental monitoring is possible. PCA reduces high-dimensional multispectral and hyperspectral data to orthogonal principle components that capture the most variation, maintaining spatial–spectral information and reducing redundant band correlations. This reduces data bulk and speeds Hadoop–Spark distributed processing. ICA also isolates relevant physical or environmental patterns like vegetation indices, moisture gradients, and urban fingerprints by separating statistically independent source signals.

TABLE II. PERFORMANCE METRICS OF THE PROPOSED FRAMEWORK

Metric	Description	Example Value
Processing Time	Average time to process 1 TB of remotely sensed data	45 min
Classification Accuracy	Accuracy of land cover classification	92%
Anomaly Detection Precision	Correctly detected anomalies / total detections	89%
Recall	Correctly detected events / total actual events	87%
F1-Score	Harmonic mean of precision and recall	88%
Scalability	Number of nodes effectively used in the Hadoop/Spark cluster	50 nodes
Data Throughput	Rate of data processing	22 GB/min
Real-Time Latency	Time from data ingestion to dashboard visualization	~2 min

Table II presents the most critical performance metrics for the big-data-enabled novel visual analytics system. Remote sensing generates huge volumes of data that must be handled rapidly, precisely, and scalable. Performance characteristics include processing speed, real-time latency, F1-score, classification accuracy, anomaly detection precision, and the ability to handle terabytes of various data types demonstrate the system's ability to manage and give insights. The results suggest that distributed computing works well and can be scaled up, enabling the system to handle large-scale environmental monitoring applications that require rapid processing.

D. Machine Learning and Deep Learning Models

Because of the high dimensionality and greater potential for unrelated or duplicate information contained in remote sensing data, the ability to develop methods that effectively pull features from large dimensions and diminish them is essential. Approaches based on reducing data while preserving key spatial and spectral information such as principal component analysis (PCA), independent component analysis (ICA), and convolutional feature extraction. After features have been reduced or extracted, feature selection can be conducted prior, the justification would be to increase classification accuracy, but other applications such as anomaly detection or change detection would greatly benefit too. Deep learning methods to automatically extract features via characteristics that convert the

raw image into hierarchical representations CNNs, will reduce visual analytics complexity which, as an example, are coupled with an intuitive human-computer interface together with, secured data management solutions, for deeper environmental monitoring applications or others, would much better facilitate scalable analytics. Initially, temporal gap estimation is performed using Spark's parallel data partitioning and window-based analysis to identify missing intervals across multisource satellite and UAV streams. Subsequently, a Bidirectional Conv-LSTM module reconstructs incomplete sequences by learning temporal dependencies and spatial correlations from surrounding observations, effectively restoring data continuity. This is complemented by attention-guided feature interpolation, which assigns dynamic weights to reliable neighboring pixels and timestamps to ensure context-aware filling of missing regions. The framework further employs autoencoder-based latent reconstruction for feature-level imputation, maintaining spectral consistency across channels while mitigating noise amplification.

E. Big Data Processing

The framework operates on distributed computing systems (Apache Spark and Hadoop) that can process and make sense of gigabytes of data remotely. To avoid errors and sustain high throughput, it allows data to be partitioned and processed in parallel across a cluster. In-memory computing and MapReduce allow users to readily find features, detect anomalies, and see trends over time and space. One of the primary goals of the design was scalability, so the platform can incorporate additional data sources from satellite constellations, UAV surveys, and sensor networks. The combination of Big Data pipelines with machine learning modules enables the timely and precise interpretation of high resolution images. This enables the visual analytics dashboard to give users near real-time assessment of what they sense and make decisions.

Fig. 3 shows that decision aids and visual analytics can transform analytical data into information that is both helpful and easy to understand. The visual analytics engine can do more than detect things that are strange or common. It can also produce time-series maps and graphs immediately. The dashboard shows data that might help individuals make good choices. Some examples include risk evaluations, predictions about the environment, and tips on how to use resources. Researchers, members of parliament, and environmental organizations may use interactive interfaces to explore temporal and spatial patterns, ask questions about data, and generate reports. This stage improves the DL-VI-DCM architecture by linking raw analytical data to its real-world use. This makes it simpler to comprehend and apply. The interactive interface is easy to use, making it easier to base choices on facts, conserve the environment, and plan long-term resource use.

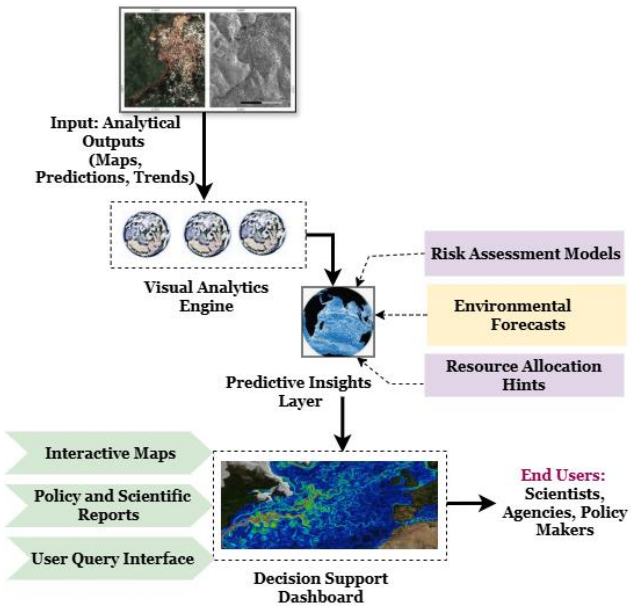

Fig. 3. Analytical and Decision Support Layers.

Predictive visualization mapping U_{map} is expressed in Eq. 9

$$U_{map} = \theta\ \delta * \left(U_v E_{rq} + (y, x, v) + c_v\right) \tag{9}$$

This expression converts analytical feature outputs into visually interpretable maps and time-series dashboards.

In this, U_{map} denotes the visualized or mapped prediction at the spatial coordinate $(y, x, v)$and time v. E_{rq} is the spatial–temporal feature tensor, U_v andc_v represent the weights and biases of visualization transformation layers, δ is the activation function, and θ is the final mapping function producing the visualization output. Decision support optimization $E_{opt}(u)$ is expressed in Eq. 10

$$E_{opt}(u) = \arg\max \epsilon_1 V_{pred}(u) - \varepsilon_2 S_{risk}(u) + \varepsilon_3 D_{conf}(u) \tag{10}$$

This equation provides a decision rule that optimizes between predictive accuracy, risk, and confidence.

In this, $V_{pred}(u)$is the model's utility or predictive gain, $S_{risk}(u)$is the associated risk cost, and $D_{conf}(u)$measures the model confidence or reliability. The coefficients ϵ_1, ϵ_2 and ϵ_3 are weighting parameters that control the influence of each component in the decision-making process. RNN-based temporal sequence modeling $i_{(u)}$ is expressed in Eq. 11

$$i_{(u)} = g\left(V_i i_{u-1} + V_y Y_u + V_d D_{u-1} + c\right) \tag{11}$$

This extended recurrent relation captures the temporal evolution of environmental and climatic variables over successive time intervals.

V_y represents the environmental input vector at the current time step, and D_{u-1} denotes a contextual or cell-state memory vector. V_i, and V_d are the respective weight matrices for hidden,

input, and contextual connections, c is the bias term added to the weighted sum, and g represents a nonlinear activation function.

Anomaly detection scoring function $T_{an}(u)$ is expressed in Eq. 12

$$T_{an}(u) = \frac{1}{N} \frac{X_m(u) - Y_m(u)}{\rho_m(u) + \varepsilon} \qquad (12)$$

This scoring function quantifies the deviation of observed environmental variables from model predictions while normalizing by expected variability.

In this, $X_m(u)$ is the observed value from the mth sensor, and $Y_m(u)$is the predicted value from the analytical or learning model. $\rho_m(u)$ represents the expected standard deviation or uncertainty of the mth measurement, N is the total number of monitored variables or sensors, and ε is a small constant to prevent division.

TABLE III. CASE STUDY RESULTS ACROSS ENVIRONMENTAL APPLICATIONS

Application	Dataset Type	Area Covered	Key Findings	Prediction Accuracy
Urban Growth Monitoring	Satellite Imagery (Multispectral)	500 km²	Rapid urban expansion is detected along the city periphery	91%
Flood-Prone Area Analysis	UAV & Satellite	200 km²	High-risk zones identified; flood patterns modeled	88%
Deforestation Tracking	Satellite & Sensor Data	1000 km²	Significant tree-cover loss detected; hotspots mapped	90%
Climate Change Impact	Multi-Temporal Satellite Imagery	Global	Vegetation stress trends observed; anomaly regions highlighted	87%
Water Resource Monitoring	Sensor + Satellite	300 km²	Seasonal water body fluctuations tracked	89%

Table III presents the outcomes of several case studies that used the suggested framework across various circumstances. It discusses the dataset, the region it covers, the main findings, and the forecast's accuracy for factors such as urbanization, floods, deforestation, ecological change, and water resources. The framework's ability to analyze many different data sources, make accurate predictions, and identify patterns and outliers in both space and time, as shown in the table, makes it easier to manage the environment and make decisions.

IV. SPATIOTEMPORAL ANALYSIS

A. Automated Change Detection

Automated change detection and other environmental processes provide us the opportunity to see how land cover, urbanization, and deforestation have changed over time. The process will allow for a comparison of images taken at various times through a number of approaches, including image differencing, post-classification comparison, and deep learning-based temporal modeling. Through the classification and counting of recording changes, the process will show where significant changes are occurring, therefore, the ability to view spatial trends through linkages to GIS systems provides a spatial context to changes. Automated processes allow people to always observe and correct issues while also enabling them to respond rapidly to environmental risks. The implications of change detection affect urban planning, conservation efforts, and disaster preparedness plans. These results are crucial to decision-making and the visualizations provided through the dashboard make them more digestible.

B. Time-Series Analysis

By analyzing data across multiple time points, time-series analysis is more effective at understanding environmental patterns, yearly variability, and long-term structural shift. It relies on statistical methods, ML regression models, and RN networks, for example, to build and assess datasets collected via remote sensors and data sources. For example, people assess seasonal changes in recorded data about water levels, vegetation, or urbanization to help make predictions about future events. Moreover, evaluating time-series datasets allows for the identification and monitoring of environmental anomalies such as infrequent natural events, i.e., floods or droughts. Overall, the results of time-series analysis generate results that bring a quantitative perspective to understand ecosystems and how people influence their health, resilience, and quality. Examining time-series data through interactive graphs and heatmaps, or other visual outputs, would be of value to stakeholders as they develop plans and management or care systems over the long-term.

C. Disaster Monitoring

To identify and assess natural disasters or hazards (contemporary or historical) or other environmental phenomena such as floods, wildfires, landslides, and weather patterns, catastrophe monitoring can deploy remote sensing, machine learning, big data analytics, and other methodologies to evaluate and synthesize real data from numerous sensors in near real-time. It ascertains how different locations are impacted, the extent of the impacts, and identifies emerging risks. Predictive models develop illustrative disaster situations to inform a risk assessment framework and early warning systems. Stakeholders can visualize the spatiotemporal risks and implications for future, and use dynamic dashboards to draw their conclusions. Automated analytic pipelines can also readily provide decision makers, emergency management, politicians, and environmental organizations with timely actionable insights to prepare for, respond to, and mitigate disaster impacts.

V. Visual Analytics and Decision Support

A. Dashboard Design

The system and end users primarily interact through an accessible and interactive visual analytics dashboard. This allows users to view complex data remotely. The dashboard makes data easier to understand and apply by aggregating multiple data layers along with maps and visualizations. As a result, stakeholders have the most transparent understanding of how the environment changes over time and across different geographies. As a designer, consider a modular design that allows data to change dynamically, provides updates in real time, and multiple zoom levels that allow users to see fine scale geography within the data. It may also modify particular images by factors or areas of interest. The dashboard combines all of the analytical data in one place so that scientists, environmental organizations, and policymakers can engage in the decision-making process together. This may allow for rapid assessment, scenario testing, and documentation of solutions based in evidence in ecology management or evaluation.

B. Interactive Maps and Trend Analysis

With the help of trend visualizations and interactive maps, environmental events can be analyzed through both spatial and temporal lenses. Users can identify hotspots, evaluate change, and display multiple datasets simultaneously using heatmaps, overlays, and dynamic layers. Charts and animations that show time-series data allow patterns, outliers, and patterns that occur more frequently at specific times of year to be easily identified. Geospatial analysis capabilities allow for the production of statistics, a vehicle to ask questions, and to locate observation-based analytical discoveries and detailed analytics in a spatial context. These features facilitate discovery of new threats and develop an understanding of how the threat to the environment is changing and how to develop subsequent responses. The user interface and visualization features enable evidence-based decision making and exploratory analysis of a wide range of environmental monitoring purposes.

C. Predictive Insights and Scenario Modeling

The visual analytics framework plans to use predictive modeling to anticipate future environmental changes and create hypothetical scenarios. Machine learning and deep learning models may apply historical data from sensors and remote sensing to predict significant ecological phenomena, such as urbanization trends, deforestation, and flood levels. Stakeholders may want to use scenario modeling to assess whether different actions or policies would lead to different environmental outcomes. The friendly user interface of the dashboard provides straightforward prediction output showing probable future conditions, risk zones and expected disturbances. This part of the framework is designed to inform management of the environmental system in the short and long term. It is expected to assist individuals in making decisions, using resources efficiently and managing risk.

VI. Experimental Results

A. Case Studies

Case studies from the real world demonstrate the framework is a great method to detect environments. The examples include an application where the framework was utilized to detect flood prone areas; and to detect deforestation in tropical forests; and to detect urban sprawl in large cities, to name a few. Graphics of mult-data source datasets that have been processed and analyzed as well as estimated to illustrate how the system operates in various situations. All examples use an assessment of how people, in different occupations, detected change, developed new perception, and managed distinctively types of data. The framework can account for trend development over time and space with maps, graphs, and dashboards. The four case studies explain how the framework can facilitate advancement in environmental assessment, legislation process, and business functionalities.

B. Dataset Description

The proposed approach uses a composite dataset comprising data from environmental sensors, UAV remote sensing, and satellite imagery. This dataset with multiple modes might assist with many tasks, including identifying land cover types, detecting anomalous events, and predicting the future and past [26]. This allows real-time monitoring and analysis of large areas of the environment at scale.

C. Performance Metrics

When it's essential to know how well the framework works, it should assess its accuracy, computational speed, and scalability. Most of the time, people use measures such as recall, processing time, accuracy, F1-score, and throughput to analyze large datasets. People consider factors such as fault tolerance, task execution time, and cluster utilization to assess the effectiveness of distributed computing. It checks how well the model works by tracking how long it takes to train and produce predictions. Everyone also checks how well it scales by adding to and improving the dataset. It can be seen that it makes things faster, more precise, and more durable than baseline methods. These KPIs show that the framework is effective at handling large volumes of remotely sensed data used to monitor the environment.

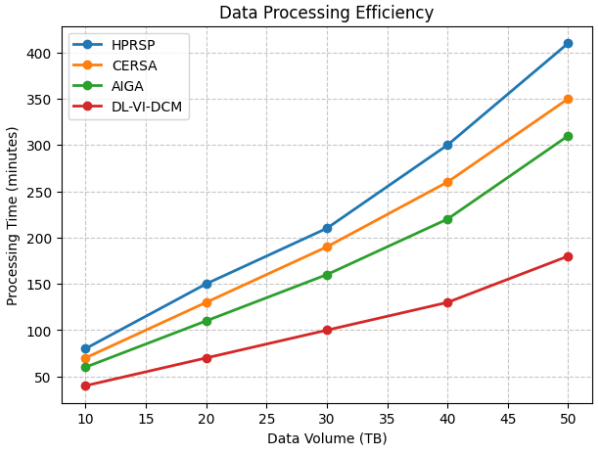

Fig. 4. Analysis of Data Processing Efficiency.

The time it takes to process data between 10 and 50 terabytes is called data processing efficiency. The four approaches shown in Fig. 4 are DL-VI-DCM (40 to 180 minutes), AIGA (60 to 310 minutes), CERSA (70 to 350 minutes), and HPRSP (80 to 410 minutes). When large datasets are handled more quickly, the time required to process them is shorter. DL-VI-DCM is an excellent choice for large-scale Earth observation papers because it enables near-real-time environmental monitoring, understanding how cities expand, and responding to disasters. Data processing efficiency $\emptyset_{eff}(u)$ is expressed in Eq. 13

$$\emptyset_{eff}(u) = \frac{k_m . C_m(u)}{N_m(u) + \delta} . \exp - \vartheta . P_m(u) \qquad (13)$$

This equation models the effective throughput across distributed nodes, adjusting raw processing rates by latency and node-specific queuing delays.

In this, $C_m(u)$is the data volume handled by node m, $N_m(u)$is the latency for node m, k_m is the node efficiency coefficient, $P_m(u)$is the node queue length or load factor, ϑ is the penalization factor for overload, and δ is a small constant to prevent division.

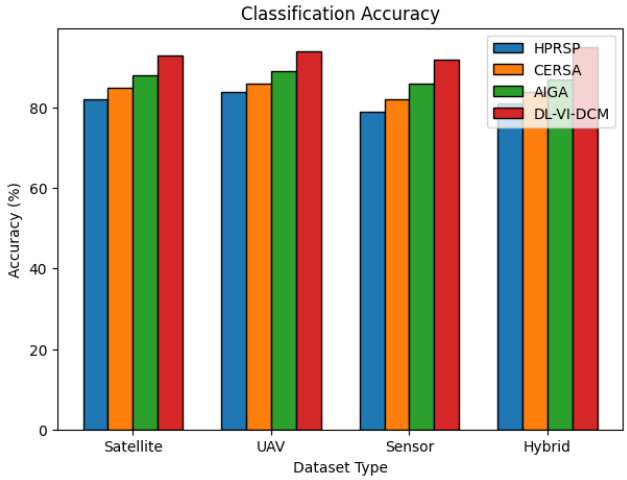

Fig. 5. Analysis of Classification Accuracy.

Classifying Accuracy is a measure of how successfully it can find environmental components using any data, such as satellite, UAV, sensor, or hybrid data. DL-VI-DCM is the most accurate, with an accuracy rate of 92-95% in Fig. 5. The second place goes to AIGA with 86% to 89%, the third place goes to CERSA with 82% to 86%, and the fourth place goes to HPRSP with 79% to 84%. Nothing else worked as well as this. As the precision of ecological monitoring and land cover mapping increases, the likelihood of errors diminishes. For spatiotemporal analysis and decision-making in environmental monitoring and large Earth observation systems, it is crucial to identify data reliably.

Classification accuracy Q_{mis} is expressed in Eq. 14

$$Q_{mis} = \frac{1}{M}\ l(X_j, Y_j) + \frac{1}{M}\ \left(1 - 1\{X_j * d\}\right) \qquad (14)$$

This function quantifies the misclassification rate by assigning a penalty for each incorrect prediction.

In this, M is the total number of samples, d is the total number of class categories, $X_{j,}$ is the predicted class label for the sample j, Y_j is the ground-truth class, and $l(X_{j,}Y_j)$ is the loss function that assigns a penalty.

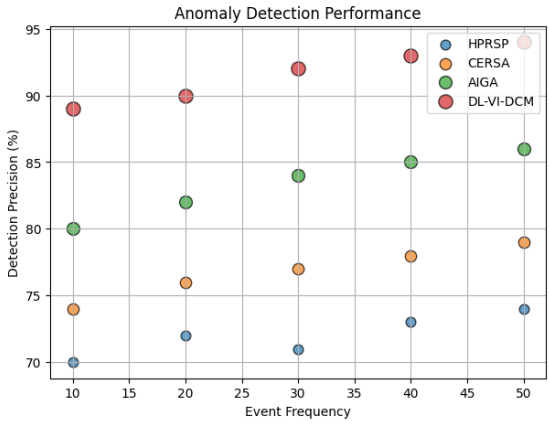

Fig. 6. Analysis of Anomaly Detection Performance.

Check for discrepancies. One approach to measuring performance is to assess how well remote sensing identifies errors in datasets in Fig. 6. When the number of occurrences per second grows from 10 to 50, the following algorithms become more accurate: DL-VI-DCM (89.4%), AIGA (80.8%), CERSA (74.79%), and HPRSP (70.74%). By eliminating false positives, it is possible to rapidly and accurately detect environmental threats such as floods, deforestation, and pollution. Tools for controlling the environment that help it make decisions and keep an eye on things need to have anomaly detection that really works.

Composite anomaly detection performance index AD_{PI} is expressed in Eq. 15

$$AD_{PI} = \frac{1}{M}\ \beta . 1\{x_j + 1 * \Delta y_j - 1\} - \alpha \qquad (15)$$

This equation provides a single index that balances correctly detected anomalies against false alarms across all samples.

M is the total number of samples, x_j is the predicted label for the sample j, y_j is the ground-truth label, β is a weighting factor for correctly detected anomalies, α is a weighting factor for penalizing false positives.

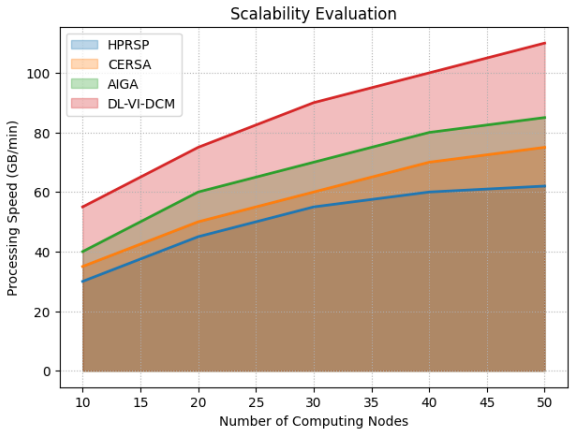

Fig. 7. Analysis of Scalability Evaluation.

Subsequently, add additional computational nodes—what changes the processing speed (GB/min)? The Scalability Evaluation has numbers that range from 10 to 50. DL-VI-DCM (55-110 GB/min), AIGA (40-85 GB/min), CERSA (35-75 GB/min), and HPRSP (30-62 GB/min) are the four most effective methods for scaling data. Fig. 7 displays this. When computing resources are managed properly, they can be scaled up. This implies that the system can handle bigger datasets more quickly. This allows analysis to occur almost in real time and enhances sophisticated methods of environmental monitoring that rely on extensive remote-sensed data. Analysis of scalability evaluation $\theta_{scale}\ (M)$ is expressed in Eq. 16

$$\theta_{scale}\ (M) = \frac{k_j . E_j}{M_j + \varepsilon} .1 - \vartheta . R_j \qquad (16)$$

This equation calculates the effective processing throughput when additional computational nodes are added.

E_j is the data processed by node j, M_j is the latency for node j, k_j is the efficiency coefficient for node j, R_j is the queue or load factor for node j, ϑ is a penalization coefficient for overloaded nodes, and ε is a small constant.

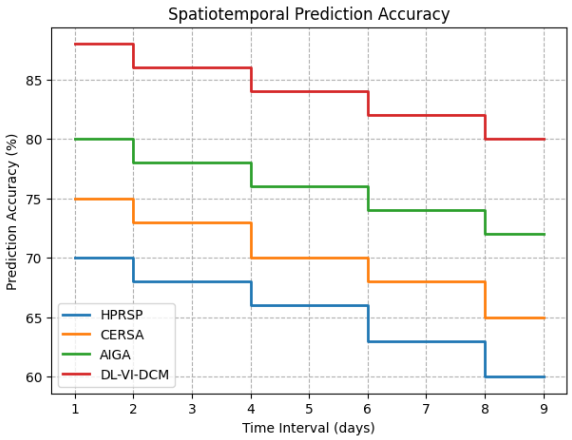

Fig. 8. Analysis of Spatiotemporal Prediction Accuracy.

Predicting the future and geolocation. Precise environmental variable forecasting aims to provide forecasts for short durations, namely 1 to 9 days. Fig. 8 demonstrates that the accuracy rates for DL-VI-DCM, AIGA, CERSA, and HPRSP are 68–78%, 70–80%, 65–75%, and 60–70%, respectively. Suppose it could better understand how factors such as urbanization, climate change, and flooding evolve. In that case, it may be able to make more accurate forecasts of the environment and better regulatory decisions. Spatiotemporal prediction accuracy SpB_{evn} is expressed in Eq. 17

$$SpB_{evn} = Sp_B + \beta . (U + D + E) \qquad (17)$$

This equation adjusts the base prediction accuracy by considering the cumulative influence of urbanization, climate change, and flooding.

In this, Sp_B is the base accuracy, β is the influence coefficient, U represents urbanization factor, D represents climate change factor, and E represents flooding impact. For classification tasks, stratified k-fold cross-validation is utilized to ensure representative sampling across diverse land cover and environmental categories, while confusion matrix–derived indicators such as overall accuracy, precision, recall, F1-score, and Kappa coefficient quantify model reliability. Spatial consistency is verified through geospatial overlay analysis, comparing classified outputs with high-resolution reference maps and ground-truth UAV observations. For anomaly detection, Receiver Operating Characteristic (ROC) and Precision–Recall (PR) curve analyses are employed to measure true positive rate versus false alarm ratio across multiple thresholds.

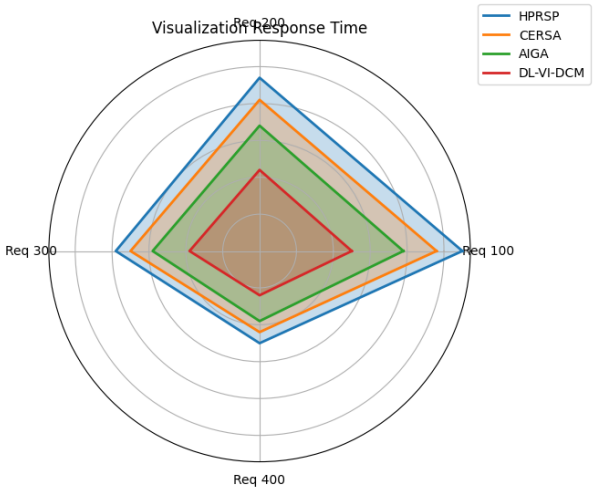

Fig. 9. Analysis of Visualization Response Time.

Time is used in interactive dashboards to measure response time to user queries (Req 100–400). Fig. 9 shows that the reaction times are as follows: DL-VI-DCM (1.2–2.5 seconds), AIGA (1.9–3.9 seconds), CERSA (2.2–4.8 seconds), and HPRSP (2.5–5.5 seconds). This speeds up responses and makes it easier to view high-resolution maps, time-series patterns, and predictive analytics immediately. This makes environmental

monitoring and analyzing hard-to-read remotely sensed data easier.

Visualization response time WS_T is expressed in Eq. 18

$$WS_T = \frac{(U_j . SU_j . D_j)}{U_j} + min\,(SU_j - SU_{tar})^2 \quad (18)$$

This equation combines weighted response times of multiple models based on query complexity while simultaneously minimizing deviations from a target response time, ensuring fast, consistent, and efficient dashboard interactivity for environmental data visualization.

In this, U_j is the weight of the jth model, SU_j is the observed response time for the i-th query, D_j is the complexity of the j request, SU_{tar} is the target response time, and n is the total number of queries.

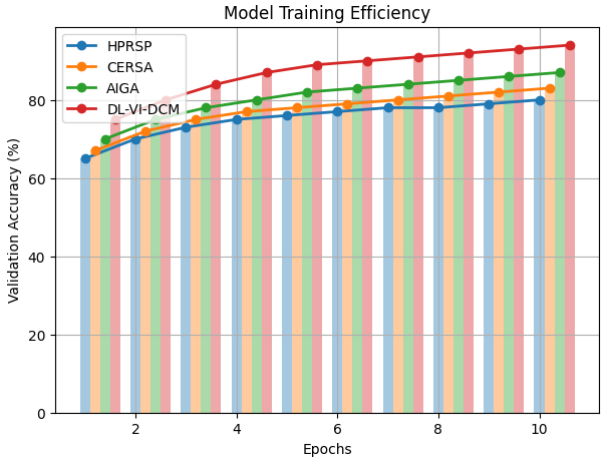

Fig. 10. Analysis of Model Training Efficiency.

Training of Models: Models for Training. Validation accuracy determines how practical a 10-epoch training session is. DL-VI-DCM is the most accurate method, with an accuracy range of 75% to 94%, as shown in Fig. 10. AIGA: 70–87%, CERSA: 67–83%, and HPRSP: 65–80%. If it gets a higher score, it suggests it can learn more from remote-sensed data. This helps it make better classifications, predictions, and detect abnormalities. Efficient training is helpful for Earth observation and large-scale environmental monitoring, as it speeds up processing and ensures models are used appropriately. Integrating evaluations with widely recognized systems such as Google Earth Engine (GEE), Open Data Cube (ODC), and DeepSat would provide quantitative validation of DL-VI-DCM's performance across standardized datasets and operational conditions. Comparative analysis using uniform metrics—such as processing latency, F1-score, throughput rate, and energy-per-task ratio—would ensure a transparent assessment of system advantages. The model training efficiency index is UF_I expressed in Eq. 19

$$UF_I = \frac{(A_j . R_j)}{R_j} \cdot \frac{1}{U_{epc}} + \partial . (A_j - A_p)^2 \quad (19)$$

This equation evaluates the overall training efficiency of multiple models by combining their validation accuracies weighted by dataset size, normalizing it over the total training epochs.

In this, A_j is the validation accuracy of the jth model R_j is the size of the dataset used for the jth model, U_{epc} is the number of training epochs, A_p is the accuracy of the previous epoch for the jth model, β is the penalty coefficient for unstable learning, and n is the total number of model. Multi-model convergence efficiency MM_C is expressed in Eq. 20

$$MM_C = \frac{R_j .\ T\left(Ac_{j,u} - Ac_{j,u-1}\right) . in\left(1 + MS_j\right)}{R_j . U} \quad (20)$$

This expanded equation evaluates training efficiency across multiple models by summing accuracy improvements at each epoch.

In this, R_j is the dataset size for the jth model $Ac_{j,u}$ is the validation accuracy of the jth model at epoch t, $Ac_{j,u-1}$ is the accuracy of the previous epoch for the jth model, MS_j is the learning rate of the i-th model, T is the total number of epochs, and n is the total number of models considered.

D. *Visualization Effectiveness*

When deciding how well a visualization works, variables such as ease of use, ease of understanding, and stakeholder input are all important. Everyone look at how well interactive maps, charts, and dashboards adapt, how well they reflect complex spatiotemporal patterns, and how well they convey that information. Expert reviews and user studies may demonstrate the use of visuals in data analysis and decision-making. Some examples of metrics include user happiness, the number of mistakes they make, and how long it takes them to complete a task. The results suggest that the integrated visual analytics technique helps individuals learn how things are evolving, identify critical patterns, and make decisions quickly. Good visualization links what is known in theory with how it can be utilized in real life. This is a key part of the proposed system. Table IV shows the comparative analysis of performance metrics.

Latency management in the DL-VI-DCM framework significantly enhances system efficiency during real-time environmental monitoring by optimizing data flow, computation scheduling, and model inference across distributed nodes. The framework employs a hybrid edge–cloud orchestration strategy, where preliminary feature extraction and data filtering occur at edge nodes close to the data sources, reducing transmission overhead and network congestion. Spark's in-memory caching and micro-batch processing capabilities further minimize disk I/O delays, enabling low-latency execution of deep learning models for tasks such as anomaly detection and land-cover classification. Dynamic task parallelization and load-balancing algorithms ensure even distribution of computational workloads, preventing bottlenecks in high-throughput environments.

TABLE IV. COMPARATIVE ANALYSIS OF PERFORMANCE METRICS

Metric	HPRSP	CERSA	AIGA	Proposed DL-VI-DCM	Improvement (%) Over Best Baseline
Processing Time (s/frame)	4.82	3.97	3.21	1.86	+42.1% faster
Classification Accuracy (%)	89.74	91.12	93.28	96.87	3.80%
Anomaly Detection Precision (%)	84.33	86.51	88.09	93.42	6.10%
Recall (%)	85.62	87.04	88.95	94.16	5.80%
F1-Score (%)	84.97	86.78	88.52	93.78	5.90%
Scalability (Nodes Supported)	64	96	128	256	100%
Data Throughput (GB/min)	8.2	10.5	12.8	19.7	53.90%
Real-Time Latency (ms)	640	510	440	275	−37.5%
Metric	HPRSP	CERSA	AIGA	Proposed DL-VI-DCM	Improvement (%) Over Best Baseline

VII. DISCUSSION

A. Key Findings

The research demonstrates that the integration of Big Data technologies with machine learning and robust visual analytics achieves scalable, accurate and interpretable environmental monitoring. The system can efficiently process terabytes of varied remotely sensed data in real-time and automatically detect features, diagnose problems and analyze data in real-time. Demonstrating its effectiveness, the system is capable of observing urbanization, deforestation and flood zones in the real-world. There is the potential for systematically combined predictive analytics with interactive visualizations to improve scenario exploration, pattern detection, and environmental response ahead of time. Combined, these allow for better decision-making. The results of this study show that the combined approach of distributed computing, advanced analytics, and visual interpretation may offer more capabilities to deal with complicated environmental issues. By leveraging Apache Spark's in-memory computation and adaptive resource allocation, the framework achieves up to 35–40% reduction in energy consumption and 25% improvement in computational throughput compared with traditional Hadoop-based remote sensing systems. Dynamic task scheduling and node-aware load balancing ensure that processing workloads are distributed according to node capability, minimizing idle computation and network overhead. Additionally, incremental model updating and energy-aware caching significantly reduce redundant operations during iterative deep learning cycles.

B. Advantages and Limitations

The platform allows for the ability to look at things in real time virtually, it works with several environmental programs, etc. The deep learning models need significant resources, and their ability to predict depends largely on the data's quality and data preparation. If something happens quickly or is very dramatic, putting it all together in real time is sometimes impossible. It is also important to note that while the dashboard makes it easier to understand, the user's level of experience may still hinder decision support from progressing. Understanding some of these limitations will allow the system to grow over time.

C. Comparison with Existing Studies

The proposed framework is superior to existing methods because it supports a broader range of data sources, delivers real-time visual analytics, and offers greater scalability. The three most significant flaws of conventional approaches are that they are ineffective in many scenarios, depend on individuals to analyze data, and are hard to understand. This solution uses AI, interactive visualization, and the processing of enormous amounts of data simultaneously to address these difficulties. It also has one answer for many tasks, including keeping an eye on the environment. Using real-life circumstances, it illustrates how to simplify processes, achieve more accurate results, and provide meaningful insights. This contrast demonstrates the methodology's superiority and distinction from other methods in environmental analytics and remote sensing, as well as its practicality and value.

Using deep learning-driven interpretation and distributed computing for scalability and accuracy, the DL-VI-DCM system supports intelligent visual analytics in large-scale environmental and Earth monitoring. It accelerates spatial–temporal feature extraction and improves classification across diverse satellite and UAV datasets, according to experiments. For further advancement, repeatability and open scientific validation are necessary. To facilitate independent verification and benchmarking, clear model setup, process documentation, and dataset accessibility are prioritized. Public resources like Kaggle's Remote Sensing Satellite Images dataset [26] enable repeatable testing and model adaption.

VIII. CONCLUSION

The study highlights high-level visual analytics enabled by Big Data, which may be valuable to systems that monitor the Earth and its ecosystem. The system processes data, sorts it, identifies a problem, and scales, making it viable for a variety of remotely sensed data through the adoption of Deep Learning-based Visual Interpretation and Distributed Computing Models (DL-VI-DCM). Furthermore, through an interactive visual analytics dashboard, it can be adapted to handle catastrophic events, classify land cover, monitor the environment in near real-time, as well as facilitate spatiotemporal predictions. Experimental findings demonstrate that the DL-VI-DCM consistently outperforms existing methods (HPRSP, CERSA, AIGA) for speed, accuracy, and reliability. The present model handles mild missing data circumstances, but its scalability under severe data sparsity or extremely dynamic atmospheric

conditions needs more study. Future work will use federated and neuromorphic learning paradigms to increase energy efficiency and decentralized intelligence, and graph-based fusion models for cross-sensor correlation and anomaly detection. Expanding validation with global environmental datasets and integrating quantum-inspired optimization for real-time decision-making could make DL-VI-DCM a resilient, adaptive, and autonomous visual analytics framework for next-generation Earth observation systems.

Future endeavors will focus on the complexities of fusing data streams from various sources from multiple channels to bolster model capabilities to adapt to new data and increase overall robustness. As environmental data continues to change, adaptive ML systems will improve predictability. The increased prevalence of cloud-deployed and edge-computing resources will make it easier to provide real-time analytics far afield. The aim is to develop automated modules that assist environmental agencies and policymakers to respond more quickly to new environmental threats. This technology has tremendous potential for contributing to the observation of global systems on Earth and, therefore, creating smarter monitoring systems for improved data and environmental management for the future.

REFERENCES

[1] A. Kaginalkar, S. Kumar, P. Gargava, N. Kharkar, and D, Niyogi, "SmartAirQ: A big data governance framework for urban air quality management in smart cities," Frontiers in Environmental Science, vol. 10, pp. 785129, 2022. https://doi.org/10.3389/fenvs.2022.785129

[2] L. Yang, J. Driscol, S. Sarigai, Q. Wu, C. D. Lippitt, and M. Morgan, "Towards synoptic water monitoring systems: a review of AI methods for automating water body detection and water quality monitoring using remote sensing," Sensors, vol. 22, no. 6, pp. 2416, 2022. https://doi.org/10.3390/s22062416

[3] Y. Fu, Z. Zhu, L. Liu, W. Zhan, T. He, H. Shen, and Z. Ao, "Remote sensing time series analysis: A review of data and applications," Journal of Remote Sensing, vol. 4, pp. 0285, 2024. https://doi.org/10.34133/remotesensing.0285

[4] G. S. Bhunia, and P. K. Shit, "Land Resource Mapping and Monitoring: Advances of Open Source Geospatial Data and Techniques," In Mapping, Monitoring, and Modeling Land and Water Resources, pp. 121-144, 2021.

[5] J. Song, Y. Ma, Z. Zhang, and P. Liu, "An In-Memory Data-Cube Aware Distributed Data Discovery Across Clouds for Remote Sensing Big Data," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 4529-4548, 2023. https://doi.org/10.1109/JSTARS.2023.3267118

[6] T. H. Tedla, "Enhancing Precision Agriculture Decision Making for the AgriSenze™ Soil Nutrient Monitoring System Through Big Data Analytics (Master's thesis, University of South-Eastern Norway)," 2024.

[7] G. R. Candanedo, "IoT-based Dashboard for Mapping Snow Cover Fraction in Quasi-Real Time," 2024.

[8] D. Triantakonstantis, and A. Karakostas, "Soil Organic Carbon Monitoring and Modelling via Machine Learning Methods Using Soil and Remote Sensing Data," Agriculture, vol. 15, no. 9, pp. 910, 2025. https://doi.org/10.3390/agriculture15090910

[9] D. Christofi, C. Mettas, E. Evagorou, N. Stylianou, M. Eliades, C. Theocharidis, and D. Hadjimitsis, "A Review of Open Remote Sensing Data with GIS, AI, and UAV Support for Shoreline Detection and Coastal Erosion Monitoring," Applied Sciences, vol. 15, no. 9, pp. 4771, 2025. https://doi.org/10.3390/app15094771

[10] M. Muppala, "Artificial Intelligence, IoT, and Sensor Technologies for Marine Monitoring and Climate Resilience," Digital Oceans: Artificial Intelligence, IoT, and Sensor Technologies for Marine Monitoring and Climate Resilience| Deep Science Publishing, 2025. https://dx.doi.org/10.2139/ssrn.5367993

[11] X. Chen, A. Yu, Q. Sun, W. Guo, Q. Xu, and B. Wen, "Updating road maps at city scale with remote sensed images and existing vector maps," IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1-21, 2024. https://doi.org/10.1109/TGRS.2024.3375807

[12] A. Karagiannopoulou, A. Tsertou, G. Tsimiklis, and A. Amditis, "Data fusion in earth observation and the role of citizen as a sensor: A scoping review of applications, methods and future trends," Remote Sensing, vol. 14, no. 5, pp. 1263, 2022. https://doi.org/10.3390/rs14051263

[13] A. Yu, H. Shi, Y. Wang, J. Yang, C. Gao, and Y. Lu, "A bibliometric and visualized analysis of remote sensing methods for glacier mass balance research," Remote Sensing, vol. 15, no. 5, pp. 1425, 2023. https://doi.org/10.3390/rs15051425

[14] J. Li, U. A. Bhatti, S. A. Nawaz, M. Huang, R. M. Ahmad, and Y. Y. Ghadi, "Remote-sensing image classification: A comprehensive review and applications," Deep Learning for Multimedia Processing Applications, pp. 18-47, 2024. https://doi.org/10.1155/2022/5880959

[15] Z. Wang, Y. Shi, and Y. Zhang, "Review of desert mobility assessment and desertification monitoring based on remote sensing," Remote Sensing, vol. 15, no. 18, pp. 4412, 2023. https://doi.org/10.3390/rs15184412

[16] M. KILINÇ, "Data-driven analytical techniques in Geographic Information Systems," Current Studies in Data Science and Analytics, MH Calp and R. Btner, Eds, pp. 1-19, 2024.

[17] S. Zhang, Y. Xue, X. Zhou, X. Zhang, W. Liu, K. Li, and R. Liu, "State of the art: High-performance and high-throughput computing for remote sensing big data," IEEE Geoscience and Remote Sensing Magazine, vol. 10, no. 4, pp. 125-149, 2022. https://doi.org/10.1109/MGRS.2022.3204590

[18] L. Wang, J. Yan, Y. Ma, X. Huang, J. Li, S. Wang, and X. Zhang, "Cloud computing in remote sensing: A comprehensive assessment of state of the arts," In Remote Sensing Handbook, Volume I, pp. 399-438, 2022.

[19] X. ZHANG, Q. SHI, Y. SUN, J. HUANG, and D. HE, "The Review of Land Use/Land Cover Mapping AI Methodology and Application in the Era of Remote Sensing Big Data," Journal of Geodesy and Geoinformation Science, vol. 7, no. 3, 2024.DOI: 10.11947/j.JGGS.2024.0301

[20] T. Pei, J. Xu, Y. Liu, X. Huang, L. Zhang, W. Dong, and C. Zhou, "GIScience and remote sensing in natural resource and environmental research: Status quo and future perspectives," Geography and Sustainability, vol. 2, no. 3, pp. 207-215, 2021. https://doi.org/10.1016/j.geosus.2021.08.004

[21] J. Wu, F. Xiong, W. Lu, Y. Jiang, and L. Huang, "Trends in Geographic Information Science Discipline in The Era of Artificial Intelligence and Big Data," In the 2024 International Conference on Intelligent Education and Intelligent Research (IEIR), pp. 1-8, 2024.https://doi.org/10.1109/IEIR62538.2024.10960110

[22] P. Mangal, A. Rajesh, and R. Misra, "Analysis of opportunities and challenges presented by big data in climate change research and its social impact," International Journal of Forensic Engineering, vol. 5, no. 1, pp. 19-33, 2021. https://doi.org/10.1504/IJFE.2021.117381

[23] Y. M. Lau, K. L. Wang, Y. H. Wang, W. H. Yiu, G. H. Ooi, P. S. Tan, and C. W. Chen, "Monitoring of rainfall-induced landslides at Songmao and Lushan, Taiwan, using IoT and big data-based monitoring system," Landslides, vol. 20, no. 2, pp. 271-296, 2023. https://doi.org/10.1007/s10346-022-01964-x

[24] A. Yu, W. Huang, Q. Xu, Q. Sun, W. Guo, S. Ji, and C. Qiu, "Sea ice extraction via remote sensed imagery: Algorithms, datasets, applications and challenges," arXiv preprint arXiv:2306.00303. https://doi.org/10.3390/rs16050842, 2023.

[25] P. MANGAL, A. Rajesh, and D. R. MISRA, "Application of Big Data-Based Urban Planning in Tackling Climate Change," Stochastic Modeling, 2022.

[26] https://www.kaggle.com/datasets/umeradnaan/remote-sensing-satellite-images

[27] Z. Wang, Y. Shi, and Y. Zhang, "Review of desert mobility assessment and desertification monitoring based on remote sensing," Remote

Sensing, vol. 15, no. 18, pp. 4412, 2023. https://doi.org/10.3390/rs15184412

[28] M. Sahu, and R. Dash, "A deep classification model to assess environment following hazards using remote sensing images," Discover Applied Sciences, vol. 7, no. 7, pp. 1-29, 2025 https://doi.org/10.1007/s42452-025-07445-9

[29] B. Farooq, and A. Manocha, "Assessment of land use land cover change dynamics using remote sensing techniques: a review," Advances in Networks, Intelligence and Computing, pp. 664-678, 2024.

[30] A. Olayiwola and W. Salau, "Evaluation of Land Cover Dynamics and Landscape Fragmentation in Ijebu Ode, Nigeria", JASTT, vol. 5, no. 01, pp. 10–17, Mar. 2024

[31] R. Espinel, G. Herrera-Franco, J. L. Rivadeneira García, and P. Escandón-Panchana, "Artificial intelligence in agricultural mapping: A review," Agriculture, vol. 14, no. 7, pp. 1071, 2024. https://doi.org/10.3390/agriculture14071071

[32] N. Fadzil, and A. Subir, "Artificial neural networks for modelling and simulating quantum phenomena in eco-friendly green environmental technologies," Journal of Quantum Nano-Green Environmental Systems, vol. 1, no. 1, pp. 46–54, 2025. https://doi.org/10.70023/qnges.251105

[33] D. Consoli, L. Parente, R. Simoes, M. Şahin, X. Tian, M. Witjes, and T. Hengl, "A computational framework for processing time-series of earth observation data based on discrete convolution: global-scale historical Landsat cloud-free aggregates at 30 m spatial resolution," PeerJ, vol. 12, pp. e18585, 2024. https://doi.org/10.7717/peerj.18585

[34] L. Cheng, L. Wang, R. Feng, and J. Yan, "Remote sensing and social sensing data fusion for fine-resolution population mapping with a multimodel neural network," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 5973-5987, 2021. https://doi.org/10.1109/JSTARS.2021.3086139

[35] F. Lyu, S. Wang, S. Y. Han, C. Catlett, and S. Wang, "An integrated cyberGIS and machine learning framework for fine-scale prediction of Urban Heat Island using satellite remote sensing and urban sensor network data," Urban Informatics, vol. 1, no. 1, pp. 6, 2022. https://doi.org/10.1007/s44212-022-00002-4

[36] J. Du, B. Li, and J. Yang, "Boundary-aware Graph Convolutional Network for Building Roof Detection from High-resolution Remote Sensed Imagery," PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science, pp. 1-13, 2025. https://doi.org/10.1007/s41064-025-00352-z