Journal of Geography and Cartography 2025, 8(1), 10224.
https://doi.org/10.24294/jgc10224

Article

Harnessing artificial intelligence (AI) towards the landscape of big earth data: Methods, challenges, opportunities, future directions

Zarif Bin Akhtar[1,*], Ahmed Tajbiul Rawol[2]

[1] Department of Computing, Institute of Electrical and Electronics Engineers (IEEE), Piscataway, NJ 08854, USA
[2] Department of Computer Science, Faculty of Science and Technology, American International University-Bangladesh (AIUB), Dhaka 1229, Bangladesh
*** Corresponding author:** Zarif Bin Akhtar, zarifbinakhtarg@gmail.com

CITATION

Akhtar ZB, Rawol AT. Harnessing artificial intelligence (AI) towards the landscape of big earth data: Methods, challenges, opportunities, future directions. Journal of Geography and Cartography. 2025; 8(1): 10224. https://doi.org/10.24294/jgc10224

ARTICLE INFO

Received: 30 November 2024
Accepted: 31 December 2024
Available online: 21 January 2025

Abstract: The integration of Big Earth Data and Artificial Intelligence (AI) has revolutionized geological and mineral mapping by delivering enhanced accuracy, efficiency, and scalability in analyzing large-scale remote sensing datasets. This study appraisals the application of advanced AI techniques, including machine learning and deep learning models such as Convolutional Neural Networks (CNNs), to multispectral and hyperspectral data for the identification and classification of geological formations and mineral deposits. The manuscript provides a critical analysis of AI's capabilities, emphasizing its current significance and potential as demonstrated by organizations like NASA in managing complex geospatial datasets. A detailed examination of selected AI methodologies, criteria for case selection, and ethical and social impacts enriches the discussion, addressing gaps in the responsible application of AI in geosciences. The findings highlight notable improvements in detecting complex spatial patterns and subtle spectral signatures, advancing the generation of precise geological maps. Quantitative analyses compare AI-driven approaches with traditional techniques, underscoring their superiority in performance metrics such as accuracy and computational efficiency. The study also proposes solutions to challenges such as data quality, model transparency, and computational demands. By integrating enhanced visual aids and practical case studies, the research underscores its innovations in algorithmic breakthroughs and geospatial data integration. These contributions advance the growing body of knowledge in Big Earth Data and geosciences, setting a foundation for responsible, equitable, and impactful future applications of AI in geological and mineral mapping.

Keywords: artificial intelligence (AI); big earth data; computer vision; data science; deep learning (DL); earth observations; geospatial data; machine learning (ML)

1. Introduction

In an era defined by rapid technological advancements, the convergence of Big Earth Data and Artificial Intelligence (AI) is revolutionizing geological and mineral mapping, offering transformative potential for the geosciences and resource management sectors.

Big Earth Data, characterized by its vast, multi-dimensional, and complex nature, provides unparalleled opportunities for analyzing Earth's surface and subsurface dynamics [1–3]. When coupled with AI and machine learning techniques, this data can be translated into actionable insights, facilitating deeper understanding of geological formations, mineral distributions, and critical Earth processes. Traditional geological mapping approaches, while reliable, have relied on labor-intensive methods such as manual interpretation and extensive fieldwork.

These processes, though effective, are often time-consuming, expensive, and constrained in scope [4–6]. The integration of AI into remote sensing and Big Earth Data analysis marks a paradigm shift, enabling the development of efficient, accurate, and scalable mapping techniques.

AI-powered models, particularly machine learning and deep learning algorithms, can identify subtle patterns and anomalies within massive datasets—details that might otherwise go unnoticed by human analysts [7–9]. This capability significantly enhances the precision and reliability of geological interpretations and fosters innovative solutions for industries such as mining, environmental management, and natural resource exploration.

This study provides a comprehensive analysis of the integration of AI with Big Earth Data in geological and mineral mapping. It delves into the critical criteria for selecting AI methodologies, evaluates their performance in comparison to traditional techniques, and examines the ethical and social implications of their application [10–12]. Notably, it highlights the role of organizations like NASA in advancing the use of AI for geospatial data analysis, demonstrating the current relevance and future potential of these technologies. Additionally, the paper incorporates case studies and quantitative analyses to showcase practical applications and performance metrics, emphasizing innovation in algorithms and integration methods.

By addressing challenges such as data quality, computational demands, and model transparency, this research aims to present actionable insights for responsible and equitable applications of AI in geosciences. Through a balanced discussion of the opportunities, challenges, and ethical considerations, this study seeks to advance the understanding of how AI and Big Earth Data can contribute to more sustainable, efficient, and informed approaches to geological exploration and resource management.

2. Methods and experimental analysis

This study adopts a multi-faceted approach combining Big Earth Data, machine learning algorithms, and remote sensing technologies to advance geological and mineral mapping. The methodology is systematically structured into the following phases: data acquisition, data preprocessing, model development, validation, and application to case studies. To ensure clarity and rigor, the selected AI techniques and case studies are elaborated upon, emphasizing their innovative contributions and quantitative evaluation metrics.

The data acquisition phase involves collecting extensive Big Earth Data from diverse sources, including satellite imagery, airborne sensors, and geospatial databases. These datasets include multispectral and hyperspectral images, digital elevation models (DEMs), geological surveys, and other relevant geospatial data. Publicly available datasets from institutions like the United States Geological Survey (USGS), European Space Agency (ESA), and Copernicus Open Access Hub are prioritized to ensure comprehensive spatial and temporal coverage. The selection criteria for these datasets focus on their relevance to the geological formations and mineral deposits under study, aligning with the research’s specific objectives and case study requirements.

For example, regions with active exploration and significant geological complexity were prioritized to demonstrate the robustness of the methodology. Given the complexity and heterogeneity of Big Earth Data, data preprocessing is a critical component of the methodology. Preprocessing tasks include data cleaning to remove noise and artifacts, normalization to standardize data ranges, and transformation to enhance data interpretability. Techniques such as median filtering and principal component analysis (PCA) are applied to improve data quality and reduce redundancy.

For multispectral and hyperspectral datasets, band selection and dimensionality reduction are conducted, focusing on spectral bands most relevant to mineral detection. Geospatial data undergo re-projection and resampling to ensure uniformity across datasets, and diverse sources are integrated through spatial alignment and temporal synchronization. This unified and high-quality dataset forms the foundation for subsequent AI-driven analysis. The model development phase involves tailoring machine learning algorithms for geological and mineral mapping. Both supervised and unsupervised learning methods are employed to address different analytical needs. Supervised learning algorithms, such as Random Forest (RF) and Support Vector Machines (SVM), are used for classification tasks, while unsupervised methods like k-means clustering and self-organizing maps (SOM) are utilized for pattern detection and anomaly identification. For more complex feature extraction and spatial pattern recognition, deep learning models, particularly Convolutional Neural Networks (CNNs), are deployed. The models are trained using labeled datasets where ground truth data, such as known mineral deposits and geological features, provide validation. The training process includes iterative optimization of model parameters to achieve high accuracy and generalization capabilities.

To address challenges related to model interpretability and performance, the methodology incorporates quantitative analysis of model outputs. Metrics such as accuracy, precision, recall, and F1-score are calculated to evaluate model performance. Additionally, confusion matrices and receiver operating characteristic (ROC) curves are used to assess classification reliability. Comparative analyses with traditional geological mapping methods, such as manual interpretation and field surveys, highlight the advancements achieved through AI integration.

The validation process employs cross-validation techniques, where the dataset is partitioned into training and testing subsets to minimize overfitting and enhance reliability. Independent datasets from different geographical regions are used to further validate the models, demonstrating their robustness across varying geological contexts. Quantitative performance comparisons with traditional methods are conducted to underscore the improvements in efficiency and precision.

Finally, the application to real-world case studies provides practical insights into the methodology's effectiveness. Selected regions with known geological complexities and significant mineral potential, such as areas with diverse lithological compositions or active mining zones, serve as test cases. The AI-driven models analyze these regions to identify mineral-rich zones and geological features, with results compared to existing geological maps and field data. This phase includes detailed quantitative performance metrics and visual representation of results through enhanced charts, graphs, and geospatial visualizations to effectively convey complex findings.

The methodology also integrates a discussion of the ethical and social considerations, addressing issues like data privacy, environmental sustainability, and equitable access to AI technologies in geological exploration. By highlighting these dimensions, the research emphasizes responsible innovation and the need for balanced technological advancements. Through this comprehensive and systematic methodology, the study not only demonstrates the feasibility of integrating Big Earth Data and AI for geological and mineral mapping but also establishes a framework that is replicable and adaptable for future research in geosciences. The focus on quantitative analysis, innovative algorithms, and practical applications ensures a credible and impactful contribution to the field.

3. Background research and investigative exploration towards available knowledge

Earth observation (EO) involves the interconnected systematic collection of data regarding both the physical, chemical, and also the biological systems of the Earth. This process can be further executed through various remote-sensing technologies, including satellites that constantly orbit the Earth, as well as all the associated direct-contact sensors which are located on ground-based or airborne platforms, such as the weather stations and various types of balloons. The overall information gathered through EO is very crucial for monitoring and assessing all the changes in both the natural and the built environments [1–3]. The term “Earth observation” has many types of different connotations depending on the particular region. In the areas of Europe, it often refers specifically to the satellite-based remote sensing, though it can also include in situ and the airborne observations. In the areas of the United States, the term “remote sensing” has been in use since the early 1960s and broadly refers to any observation method that utilizes types of remote sensing technology, whether it be from space, air, or any other type of ground-based platforms [4–6].

Recently, the acronym “Satellite Remote Sensing” (SRS) has begun appearing in literature as a more precise term for satellite-based observations. EO encompasses a wide array of activities, ranging from numerical measurements taken by instruments like thermometers and seismometers to photos and radar images captured by satellites or ground-based sensors. The data collected through these various means can be processed into decision-support tools such as maps and models, which are invaluable in a multitude of applications [7–9]. These include numerous weather forecasting, tracking inclusion for biodiversity and wildlife trends, measuring types of land-use changes like deforestation, and also heavily monitoring natural disasters such as fires, floods, and earthquakes.

The field of Earth observation is rapidly evolving at a high pace, with the continuous advancements in both the quality and quantity of the many types of data collected. The deployment of new remote-sensing satellites, along with their associated increasingly sophisticated in situ instruments located on the ground, in the air, and in water bodies, has resulted in comprehensive, nearly real-time observations. These technological advancements have become increasingly very much important in light of the significant impact modern human civilization has on the planet in terms of digital computing.

EO plays a critical role in mitigating these effects, such as monitoring geohazards, and offers opportunities to enhance social and economic well-being. Earth observation is a broad and multifaceted field that combines various technologies and methods to monitor the Earth's systems. Its applications are diverse and very much essential for understanding and responding to the rapid environmental changes, managing natural resources, and improving overall societal welfare [10–12].

A Digital Elevation Model (DEM) is basically a 3D computer graphics representation of the elevation data used to depict terrain or surface features of planets, moons, or asteroids. DEMs are more extensively used in terms of geographic information systems (GIS) as the foundation for digitally produced relief maps [13–15]. The term "DEM" is also often used many times interchangeably with Digital Surface Models (DSM) and Digital Terrain Models (DTM).

While DSMs include natural and man-made features like tree canopies and buildings, DTMs focus solely on representing the bare ground surface, making them crucial for applications such as flood modeling, geological studies, and land-use analysis. Terminology in the field of digital elevation modeling varies.

DEM is a broad term that can encompass both DSMs and DTMs, depending on the context. DSMs mainly represent the Earth's surface, including objects like buildings and trees, while DTMs represent only the bare ground. The term DEM is often used generically, without specifying whether it refers to DSMs or DTMs. The creation of DTMs typically involves filtering out surface objects from high-resolution DSMs through a process called "bare-earth extraction".

There are a lot of different types of DEMs, which can be very much represented either as a raster grid (often referred to as a heightmap when dealing with elevation) or as a vector-based Triangular Irregular Network (TIN). DEMs can be created using various techniques, including photogrammetry, lidar, and radar. Data for DEMs is commonly gathered through remote sensing methods, although traditional land surveying can also be used, especially in areas where remote sensing is less effective. Rendering of DEM data often involves visual forms like contoured topographic maps or color-coded elevation maps [16–18].

In some cases, oblique views are created to provide a more intuitive visualization of the terrain, with techniques like "vertical exaggeration" used to highlight subtle elevation changes. However, the use of vertical exaggeration is sometimes criticized for potentially misleading viewers about the actual landscape.

Production methods for DEMs have evolved a lot over time. While the early methods mainly relied on interpolating contour maps from land surveys, modern DEMs are mainly primarily generated using remote sensing technologies such as radar and satellite imagery. For instance, interferometric synthetic aperture radar (InSAR) is a very powerful technique that allows for the creation of DEMs over large types of areas with much high resolution.

Satellite missions like SPOT, ERS, SRTM, and ASTER have provided significant contributions to the global availability of DEM data. In planetary mapping, DEMs have become invaluable tools, especially through the use of orbital altimetry. Instruments much like the Mars Orbiter Laser Altimeter (MOLA) and the Lunar Orbital Laser Altimeter (LOLA) have been very much instrumental in terms of mapping the topography of Mars and the Moon, respectively.

Accuracy is a critical aspect of DEMs, influenced by factors such as terrain roughness, sampling density, grid resolution, and the algorithms used for interpolation and terrain analysis [19–21]. DEM quality is typically assessed by comparing DEMs from different sources. High-quality DEMs are essential for accurate modeling of terrain-related phenomena. DEMs have a wide range of applications, including geomorphology, hydrology, infrastructure design, and 3D visualizations. They are used for modeling water flow, creating relief maps, planning flights, and even for precision farming. DEMs are also used in engineering, satellite navigation, and archaeology, among other fields.

Sources of DEM data vary by a great margin globally. Free global DEM datasets like FABDEM and GTOPO30 provide a very wide coverage, although the resolution and quality can vary by a great amount significantly. Higher resolution DEMs are also available from sources like the ASTER instrument and the Shuttle Radar Topography Mission (SRTM). These types of datasets are crucial for a wide range of scientific and practical applications, including both global relief modeling and terrain analysis.

Digital Elevation Models are also the essential tools in terms of various scientific and engineering fields, offering a more detailed representations of the Earth's surface and other planetary bodies.

The development and application of DEMs still continue to advance with improvements in terms of remote sensing technology and data processing techniques. Environmental data which mainly refers to the information derived from measuring environmental pressures, along with the state of the environment, and the impacts on the interconnected ecosystems [22–24].

These components are integral to the DPSIR (Drivers, Pressures, State, Impact, Response) model, commonly used in environmental science to analyze and manage environmental issues. While environmental data primarily encompasses the "P," "S," and "I" elements of this model, it excludes socio-economic data and other statistical information often associated with the "D" and "R" components. However, for a comprehensive environmental assessment, these socio-economic and statistical data are crucial, though they are typically managed by institutions outside the environmental sector, such as National Statistical Offices.

Similarly, geo-basis data, while not classified as environmental data, are essential for effective environmental policy-making and information management. Environmental data is predominantly generated by institutions engaged in executing environmental laws or conducting environmental research [25–27]. This data serves as the backbone for environmental assessments, regulatory compliance, and policy-making. The increasing significance of environmental data in various sectors has also been recognized by the financial industry.

For instance, Bloomberg L. P. has begun providing Environmental, Social, and Governance (ESG) data through its terminals, reflecting the growing demand for this information among investors. ESG data, which includes environmental data, is becoming a critical factor in investment decisions, as investors seek to align their portfolios with sustainable and socially responsible business practices. To manage the complexities of collecting, processing, and reporting environmental data, especially in compliance with legal and regulatory requirements, Environmental Data Management Systems (EDMS) are increasingly being adopted.

These systems are designed to handle the various aspects of environmental data management, such as monitoring programs, data validation, and the generation of compliance reports. The implementation of EDMS is driven by the need to ensure accurate and timely data collection, meet compliance requirements, and reduce the administrative burden associated with environmental data management [28,29].

The growing importance of ESG factors, including environmental data, is further highlighted by predictions that ESG assets under management could reach $53 trillion within the next few years, accounting for one-third of all global assets under management. This trend is driven by factors such as fee pressure, increasing regulatory demands, and the push from asset owners for investments that are not only financially profitable but also aligned with sustainable and socially just practices. As a result, environmental data is truly playing an increasingly critical role in terms of shaping the future of both the environmental policy and global investment strategies [30–32].

An Earth observation satellite, also known as an Earth remote sensing satellite, is mainly a type of satellite which is specifically designed to observe and monitor the Earth's environment from space. These satellites also serve other types of various purposes, including environmental monitoring, meteorology, cartography, and even intelligence gathering, as seen with many spy satellites [33–35]. The most common type of Earth observation satellites are the imaging satellites, which capture images of the Earth's surface, similar to the aerial photographs.

However, some other satellites perform remote sensing without producing images, such as those which are mainly involved in GNSS (Global Navigation Satellite System) radio occultation, which measures atmospheric properties.

The root history of satellite remote sensing began mainly with the launch of Sputnik 1 by the Soviet Union on 4 October 1957. Sputnik 1 sent radio signals back to Earth, which scientists actually used to study the ionosphere, marking the first instance of satellite-based remote sensing. Following this, the United States also launched its first satellite, Explorer 1, on 31 January 1958. The data from Explorer 1's radiation detector led to the root discovery of the Earth's Van Allen radiation belts. Another significant milestone was the launch of TIROS-1 on 1 April 1960, by NASA. TIROS-1 which was the first satellite to send back the television footage of weather patterns from space, laying the root foundation for modern types of weather satellites. By 2008, there were almost over 150 Earth observation satellites in orbit, collecting vast amounts of information data daily. This number grew much significantly, reaching over 950 satellites by 2021, with the majority mainly operated by the US-based company Planet Labs. Most types of Earth observation satellites operate at relatively low altitudes, generally above 500 to 600 km, to fully capture detailed images and data.

However, the lower orbits require frequent reboost of maneuvers due to the atmospheric drag. Many Earths observation satellites, including those which are mainly operated by the European Space Agency (ESA) and the UAE, utilize low Earth orbits (LEO) to provide a much better high-resolution imagery and data [36–38].

To achieve global coverage, many Earth observation satellites are placed within the polar or Sun-synchronous orbits. A polar orbit mainly allows the satellite to scan different parts and sections of the Earth with each orbit due to the Earth's rotation.

Sun-synchronous orbits additionally ensures that the satellite passes over the same spot-on Earth at the same time each day, providing a more consistent lighting conditions for observations.

In contrast with that, geostationary orbits, located at a high altitude of 36,000 kilometers, allow the satellites to remain at fixed over a specific point on the Earth's surface, providing more continuous coverage of that particular area. This type of orbit is primarily used for meteorological satellites. Earth observation satellites have also many numerous applications, including weather monitoring, environmental monitoring, and mapping. Weather satellites, for example, tracks the cloud patterns, monitor volcanic ash clouds, and observing for smoke from wildfires. Environmental satellites detect changes within vegetation, atmospheric gases, sea conditions, and ice fields, aiding within the monitoring of droughts, oil spills, and pollution. Mapping satellites, such as the Radarsat-1 and TerraSAR-X, provide detailed terrain maps using the radar technology.

International regulations that govern the use of Earth observation satellites, particularly regarding the allocation of many radio frequencies for communication between satellites and their associated ground stations. The International Telecommunication Union (ITU) mainly defines Earth exploration-satellite service as a radiocommunication service that mainly collects and distributes data related to the Earth's characteristics and its natural phenomena. These regulations ensure that satellite operations are harmonized globally, with frequency allocations managed by national administrations.

Earth observation satellites play a crucial role in monitoring and understanding the Earth's environment, providing valuable data for various scientific, environmental, and commercial applications. Their importance continues to grow as technology advances and the demand for accurate and timely environmental data increases [39,40]. Geographic data and information, also known as geospatial data, refers to any data that is implicitly or explicitly associated with a specific location on Earth.

This type of data is critical for understanding and analyzing spatial relationships and patterns. It is commonly stored in Geographic Information Systems (GIS), which are specialized systems designed to capture, store, manipulate, analyze, and manage spatial or geographic data. GIS allows for the integration of various types of geographic information, enabling users to visualize, interpret, and understand spatial relationships and trends in the data.

There are several different types of geospatial data, including vector files, raster files, geographic databases, web files, and multi-temporal data. Each type of data has its unique characteristics and uses. For instance, vector files represent geographic features as points, lines, and polygons, making them ideal for mapping boundaries, roads, and other discrete objects.

Raster files, on the other hand, consist of grid cells that store data, such as satellite imagery or elevation models, making them suitable for continuous data representation. Geographic databases are used to store and manage large volumes of geospatial data, while web files allow for the sharing and dissemination of geographic information over the internet. Multi-temporal data refers to geospatial data collected at different times, which is essential for analyzing changes over time, such as in environmental monitoring or urban development studies.

Geospatial data and information are central to the various type areas fields of study, including geocomputation, geographic information science (GIScience), geoinformatics, and also geomatics. These fields also overlap in their focus on the acquisition, analysis, and interpretation of the geographic data [41,42]. For instance, geocomputation involves the use of computational techniques to solve geographic problems, while GIScience focuses on the theoretical and scientific aspects of geographic information systems and technology. Geoinformatics and geomatics encompass a broader range of activities, including the collection, processing, and analysis of geographic data using various technologies. In addition to these core fields, geospatial data and information are also relevant to other related disciplines such as cartography, geodesy, geography, geostatistics, photogrammetry, remote sensing, spatial data analysis, surveying, and topography. Cartography is termed the art and science of map-making, geodesy deals with the measurement and representation of the Earth, and geography studies the overall physical and human features of the Earth's surface [43,44].

Geostatistics involves statistical analysis of spatial data, photogrammetry focuses on obtaining measurements from photographs, and remote sensing refers to the acquisition of information about the Earth's surface using satellite or aerial sensors. Spatial data analysis is concerned with examining spatial patterns and relationships in data, surveying involves the precise measurement of land, and topography studies the Earth's surface features. Geographic data and information are fundamental to a wide range of scientific, engineering, and planning activities. They enable the visualization and analysis of spatial patterns and relationships, which are essential for making informed decisions in areas such as urban planning, environmental management, transportation, and disaster response. The growing availability of geospatial technologies and data has expanded the possibilities for research and application, making geographic information a vital component of modern science and technology.

The Earth Observing System Data and Information System (EOSDIS) is a very vital component of NASA's Earth Science Data Systems Program. Designed and mainly maintained by Raytheon Intelligence & Space, EOSDIS provides a very detailed with a comprehensive platform for managing and disseminating Earth science data [33–44]. This system also supports a wide range of users, from casual individuals to the specialized research scientists selected through NASA's peer-reviewed processes. EOSDIS offers many several key services, including the user support, data archiving, management, distribution, information management, and product generation, all which are overseen by the Earth Science Data and Information System (ESDIS) Project. EOSDIS is integral to the handling of data information from NASA's Earth-observing satellites. The whole system ingests, processes, archives, and distributes vast amounts of data information from these satellites, as well as data information from aircraft, field measurements, and other types of programs. The system also provides end-to-end capabilities for managing this type of data, ensuring that it is also very much accessible to a global user base. EOSDIS's capabilities are divided into many mission operations, managed by the Earth Science Mission Operations (ESMO) Project, and science operations, overseen by the ESDIS Project. These operations also include data capture, initial processing, and higher-level science data product generation, archiving, and distribution.

A key feature of the EOSDIS is its distributed system of processing the facilities and Distributed Active Archive Centers (DAACs), spread all across the United States. These DAACs serve as the custodians of Earth science data, ensuring its associated long-term preservation and accessibility. Each DAAC specializes in a very specific Earth science discipline, providing tailored services and tools to its designated user community.

The DAACs manage a massive and ever-growing database, with EOSDIS reporting around 10 petabytes of data by 2012, with a daily ingestion rate of approximately around 8.5 terabytes. EOSDIS also includes many systems for searching and accessing data, such as the Global Change Master Directory (GCMD) and the Common Metadata Repository (CMR). The GCMD is a directory of over 30,000 Earth science data information sets and services, while the CMR serves as a metadata catalog and complex registry for NASA's EOS data. In 2018, Earthdata Search replaced Reverb as the web-based client for discovering and ordering data across EOSDIS's holdings, allowing many users to search, retrieve, and order data through a much better user-friendly interface. The root level history of EOSDIS dates back to the early 1980s, when NASA began mainly exploring the feasibility of publicly accessible electronic data information systems. By 1990, the EOS mission, which also included the NASA Earth Science Enterprise, had been approved by the Congress. This mission supported the entire development of EOSDIS, designed as a long-term data and information system accessible to both the scientific community and the interconnected broader public. Over the years, EOSDIS has evolved to a great scale in terms to meet the growing demands of Earth science research, providing a more critical support for NASA's Earth-observing missions and serving a diverse global community of users around the globe.

Microsoft AI for Earth is another significant initiative launched within July 2017, focused on leveraging artificial intelligence (AI) to address critical environmental challenges. The project is part of Microsoft's broader commitment to social good, particularly in areas related to agriculture, water, biodiversity, and climate change [40–45]. The AI for Earth program is active in 40 countries, working on various projects aimed at improving the sustainability and management of the planet's resources. The initiative was launched with an initial investment of $2 million, but due to its growing impact and potential, Microsoft later expanded its strategic approach and allocated a $50 million budget to support its goals. AI for Earth has formed 50 partnerships and supported 950 projects globally, demonstrating its expansive reach and commitment to addressing environmental issues through AI-driven solutions.

One of the program's most notable developments is the creation of the "Planetary Computer." This platform also offers a wide range of tools and resources, including APIs, data catalogs through Azure storage, and many open-source tools, all designed to empower researchers and organizations to analyze and act on environmental data. The Planetary Computer represents a momentous advancement in the availability of data and computational resources for environmental science, enabling more operative and scalable solutions to the challenges facing our planet. Key figures involved in the AI for Earth initiative include Lucas Joppa, Bruno Sánchez-Andrade Nuño, Alma Cárdenas, and Harry Shum, who have been instrumental in driving the program's vision and execution.

The initiative aligns with Microsoft's broader mission of using technology for social good, with AI for Earth serving as a prime example of how AI can be harnessed to create a positive impact on both the environment and society.

4. Big earth data: The AI perspectives

Artificial Intelligence (AI) has truly become an integral part of NASA's efforts to enhance the analysis and utilization of the overall vast amounts of Earth observation data. AI, particularly through machine learning (ML), enables machines to simulate the human decision-making processes and identify many complex patterns within large amounts of datasets that would be difficult, if not impossible, for humans to discern manually. This advanced capability is especially valuable in processing and analyzing Big Data collections, such as the overall extensive data generated by NASA's Earth observing missions.

NASA's Earth Science Data Systems (ESDS) Program is also very much deeply committed to incorporating AI and ML into its associated operations to maximize the scientific return of its missions. This commitment is very much evident in the work conducted by NASA's Interagency Implementation and Advanced Concepts Team (IMPACT) at the Marshall Space Flight Center in Huntsville, Alabama. The IMPACT team comprises machine learning (ML) specialists, computer scientists, and Earth science data experts who also collaborate to develop tools and pipelines that apply within ML algorithms to NASA's Earth science datasets.

These tools significantly enhance data discovery and the overall efficiency of all the research processes. In addition to this groundbreaking work of IMPACT, AI and ML are also being utilized at NASA's Distributed Active Archive Centers (DAACs). For instance, the Goddard Earth Sciences Data and Information Services Center (GES DISC) is implementing a machine learning (ML) framework that uses natural language processing (NLP) to streamline the whole search process for data users, making it much easier for them to actually locate all the required relevant datasets.

NASA also fosters AI and ML research through its Advancing Collaborative Connections for Earth System Science (ACCESS) program. This competitive program mainly focuses on developing and implementing new technologies to manage, discover, and utilize NASA's extensive archive of Earth observations. The ACCESS 2019 solicitation specifically targeted technology developments in ML, including the innovative creation of new training datasets for machine learning applications further related to Earth science.

Furthermore, NASA supports AI and ML research through initiatives like the Frontier Development Lab (FDL), which mainly operates as an applied research accelerator at NASA's Ames Research Center in Silicon Valley, California. The FDL, created by NASA's Office of the Chief Technologist, collaborates with many academic institutions and Silicon Valley companies to advance AI research, further extending NASA's AI capabilities.

AI and ML are rapidly playing increasingly critical roles in NASA's Earth science endeavors, driving innovations that improve the utility and accessibility of Earth observation data information. Through collaborative efforts, research programs, and technological advancements, NASA still continues to push the boundaries of what AI and ML can achieve within the realm of Earth science, contributing to a more efficient and effective research and applications that will benefit both the scientific community and society at large.

5. NASA AI, DL, ML perspectives: Case studies analysis

5.1. Case study 1: Radiant earth

Radiant Earth has made significant strides within enhancing access to the Earth observation training data and machine learning (ML) models through the accelerated development of an open-access repository, Radiant MLHub. This initiative, part of the ACCESS-19 project, has focused on three main areas: creating a comprehensive global land cover training dataset, developing an open API for machine learning model registration and retrieval, and improving user accessibility through a Python client. One of the key achievements of this project is the production of the LandCoverNet dataset, a multi-mission global land cover training dataset.

This dataset consists of 8941 image chips, each measuring 256 × 256 pixels, derived from 300 geographically diverse tiles of Sentinel-2 imagery. These images span various many regions, including Africa, Asia, Australia and Oceania, Europe, North America, and also South America. The dataset also includes a yearly time series of matching Sentinel-1, Sentinel-2, and Landsat-8 imagery. Published in 2020, LandCoverNet has become one of Radiant MLHub's most downloaded datasets, underscoring its value to the machine learning and Earth observation communities.

In addition to the dataset, Radiant Earth has expanded the functionality of Radiant MLHub to support the publishing and retrieval of machine learning models. Users can now access a catalog of models through the STAC API, complete with documentation, model weights, and code, all accessible via the web interface. This expansion significantly broadens the utility of Radiant MLHub, making it a more comprehensive resource for researchers and developers working with Earth observation data. To further simplify the process of accessing these resources, Radiant Earth developed a Python client. This client allows users to programmatically search for, access, and download machine learning training datasets and models from Radiant MLHub. The introduction of this tool has reduced the complexity of interacting with the repository, making it more user-friendly and accessible to a broader audience.

Looking ahead, Radiant Earth has begun developing the next generation of Radiant MLHub, known as Source Cooperative. This new platform is designed to support extremely large training datasets and expands the dataset publishing capabilities beyond just machine learning training data. Currently in private beta, Source Cooperative is expected to enter public beta in the third quarter of 2023, marking the next phase in Radiant Earth's mission to democratize access to high-quality Earth observation data and ML resources.

5.2. Case study 2: The ESDS program project

The ESDS program project focuses on the development of an advanced platform that mainly integrates satellite observations, 3D radiative transfer simulations, deep learning (DL) techniques, and cloud computing to further enhance global cloud property retrieval. By utilizing data information from the Visible Infrared Imaging Radiometer Suite (VIIRS) on the Suomi National Polar-orbiting Satellite (Suomi NPP) and the Advanced Baseline Imager (ABI) on the Geostationary Operational Environmental Satellite-16 (GOES-16), the project aims to establish a versatile framework applicable across different satellites for retrieving the cloud properties. This initiative is crucial for refining and benchmarking various algorithms used in satellite-based cloud remote sensing. The primary objectives of the project include generating high-quality cloud physics property retrievals, such as cloud masks and cloud phases, through deep learning models that can handle multi-sensor heterogeneous data.

The project seeks to produce realistic cloud microphysics and optical property retrievals, including Cloud Optical Thickness (COT) and Cloud Effective Radius (CER), using the 3D radiative transfer simulations combined with many types of deep learning (DL) models. Another key goal is to develop scalable cloud computing-based services for enhanced processing and analyzing vast amounts of data, facilitating the implementation of cloud retrieval algorithms on a global scale. Clouds, which cover about two-thirds of Earth's surface, also plays a vital role in regulating the climate and influencing the types of various environmental cycles. Given their significance, satellite-based remote sensing has truly become essential for global cloud observation. This project aligns with the priorities set within NASA's latest Decadal Survey, which also emphasizes the importance of cloud observations in Earth science missions.

Various types of satellite sensors, both active and passive, have been developed to observe and retrieve cloud properties. Active sensors, like those available on the Cloud Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) and CloudSat missions, excel within resolving the vertical location of cloud layers, especially during nighttime and mainly over polar regions. In contrast to that, passive sensors, such as MODIS, VIIRS, and ABI, offer the superior spatial sampling rates.

5.3. Case study 3: Machine learning (ML) data workflows

Machine learning (ML) has revolutionized many fields, including satellite remote sensing. In this project, ML and deep learning (DL) techniques are mainly employed to improve cloud property retrieval. High-quality training datasets are very crucial for these models, which is why the project combines data from both active and passive sensors, advances in 3D radiative transfer simulations, and deep learning methods. This approach addresses the biases and uncertainties associated with traditional 1D radiative transfer models and facilitates the development of 3D cloud property retrievals. Additionally, the project leverages cloud computing and Big Data technologies to manage and analyze the vast archive of Earth observations efficiently.

Among the project's major accomplishments are the development of scalable satellite collocation data and toolkits, including CALIPSO-VIIRS and CALIPSO-ABI collocation datasets. These tools have been approved for the New Technology Report (NTR) and Software Release Request (SRS), and are now open-source on GitHub. The project also produced 3D radiative transfer simulation data information for synthetic cloud fields, such as fractal and Large-Eddy-Simulation (LES) clouds.

Furthermore, the collaborative team developed two deep learning (DL) models for cloud property retrieval, which have demonstrated superior performance compared to existing physics-based and deep learning approaches. These models have also received NTR approvals, and their source codes will be made available as open-source software, contributing to the broader scientific community's efforts in atmospheric remote sensing.

5.4. Case study 4: Geoweaver workflow management system

GeoWeaver is an innovative workflow management system designed to enhance the productivity and collaboration of Earth scientists by integrating Python code and Shell scripts into seamless, shareable pipelines. The system addresses the need for a flexible and intuitive tool that enables researchers to efficiently manage and execute complex workflows while ensuring that these workflows are Findable, Accessible, Interoperable, and Reusable (FAIR). By providing an intuitive interface that simplifies the creation, execution, and sharing of AI workflows, GeoWeaver aims to eliminate duplicated efforts, streamline knowledge transfer, and foster collaboration among scientists with varying technical expertise. One of the primary objectives of GeoWeaver is to make AI workflows more tangible and accessible to both beginners and experienced researchers.

The platform's design allows its associated users to quickly understand and contribute to existing projects, thereby accelerating the transition from learners to contributors. GeoWeaver achieves this by decoupling workflows from datasets and computing platforms, making them clean, safe, and portable. It also records the history of code and execution logs, ensuring that every step is permanently documented, which is crucial for maintaining the integrity and reproducibility of scientific research. Key features of GeoWeaver include the ability to execute processes on any chosen host, whether locally or remotely, and to wrap entire workflows into simple, shareable zip files. These files can be easily distributed through various channels, such as Slack, email, or social media. Additionally, GeoWeaver condenses thousands of lines of code into an intuitive graph, allowing users to browse and edit workflows within a single view. This functionality not only simplifies the overall management of complex workflows but also ensures that team members are synchronized, allowing them to collaborate effectively on the same project.

Throughout its development, the GeoWeaver team has actively engaged with the Earth science community to address user feedback and promote the adoption of the platform. The team has developed the pygeoweaver library, which has been well-received by the community, particularly among those using Python. The software is currently being utilized for various AI workflows, including the Community Multiscale Air Quality (CMAQ) AI operation site, severe weather event forecasting, and ocean eddy detection. In addition to its technical achievements, GeoWeaver has made significant strides in outreach and collaboration. The team has sponsored Earth Science Information Partners (ESIP) mini grants and worked closely with domain scientists to address the challenges of implementing FAIR Earth AI workflows. One notable project involved using GeoWeaver to create a snow workflow, which was presented at the American Geophysical Union (AGU).

Moving forward, the team plans to further expand the use of GeoWeaver in operational settings, including collaborations with NASA scientists to integrate the platform into high-performance computing (HPC) and cloud environments.

The GeoWeaver team has also contributed to the broader Earth science community by sharing their experiences with NASA's Earth Science Data System Working Group (ESDSWG) and the ESIP machine learning (ML) cluster. They have drafted a comprehensive paper titled "A Review of Earth Artificial Intelligence," which has become one of the most popular papers in the journal Computers and Geosciences. This paper aims to demystify Earth AI by providing an overview of representative AI research across the types of major spheres of the Earth system. In terms of major accomplishments, the GeoWeaver team has successfully released GeoWeaver 1.0.0-rc10, which is now ready for use, along with pygeoweaver 0.6.6, available for installation via pip. The CMAQ AI operational workflow, developed using GeoWeaver, is currently running daily, showcasing the platform's capability to support continuous, reliable operations in Earth science research.

5.5. Case study 5: Passive microwave measurements from satellites

Passive microwave measurements from satellites provide invaluable data for retrieving various surface and atmospheric parameters, but their complexity has often limited their accessibility to those with specialized satellite knowledge. These measurements, which are available in raw swath formats, are challenging to align with data from other satellites or ancillary data necessary for comprehensive analysis.

To address this issue, a project has been undertaken to resample and organize these microwave data onto fixed Earth grids, significantly lowering the barriers for broader scientific use and enabling easier integration into machine learning (ML) algorithms. The main objectives of the project include resampling microwave measurements from multiple satellites onto fixed latitude/longitude and polar grids, providing a consistent set of ancillary data aligned with these grids, and offering comprehensive documentation to assist users in creating machine learning datasets from this organized data collection. By achieving these objectives, the project mainly aims to democratize the use of passive microwave data information, allowing researchers and developers to work with these data without needing deep expertise in satellite-specific formats.

The project has also successfully resampled microwave radiances measured by the Advanced Microwave Scanning Radiometer 2 (AMSR2) onto two types of Earth grids: a global 0.25-degree latitude/longitude grid and a 25 km EASE2 polar grid for the Northern Hemisphere. The resampling process utilized the Backus-Gilbert method to achieve the high accuracy, combined with 2D interpolation to precisely place the resampled footprints. These grids support two footprint sizes: 30 km and 70 km circular footprints, depending on the frequency and polarization of the measurements. The result is a set of microwave data information that is much easier to work with and more compatible with other Earth observation datasets.

In addition to the resampled microwave data, the project also provides various ancillary datasets on the same grids, resampled to match the footprint sizes and shapes of the microwave measurements.

These ancillary datasets include land/water fraction from MODIS, precipitation data from the Integrated Multi-satellite Retrievals for GPM (IMERG), and several atmospheric parameters from the European Centre for Medium-Range Weather Forecasts Reanalysis v5 (ERA5), such as skin temperature, total column water vapor, total column cloud water, and vector winds.

These ancillary datasets are very much critical for further developing and testing retrieval algorithms, as they can serve as both the input parameters and target outputs for different ML models. The project has made many significant strides in improving the usability of this data information collection for algorithm development. The team has developed a Jupyter notebook that demonstrates how to construct a machine learning dataset using the resampled microwave data and ancillary datasets. This notebook provides simple examples that guide users through the process, making it easier for them to apply these data in their own research and development projects.

The major accomplishments of the project include the resampling of AMSR2 measurements from 2012 to 2021 onto regular grids, the collocation of ancillary data in time and space with the microwave measurements, and the development of user-friendly resources like the Jupyter notebook. These achievements represent a great significant step forward in terms of making passive microwave data more accessible and usable for a wide range of Earth science applications, particularly those involving machine learning (ML) and data integration from multiple sources.

5.6. Case study 6: DL, ML datasets

The collection of high-quality training data information is still a very significant challenge in large-area land cover classification and disturbance mapping, particularly when it comes to ensuring the minimal error and achieving higher spatial resolution than the satellite data being classified or validated. Recent advancements in deep learning (DL) and active learning approaches, combined with the availability of commercial high spatial resolution data (less than 10 meters), offer promising opportunities for generating many types of training datasets that are suitable for application to 30-meter Landsat and 10 to 20-meter Sentinel-2 data.

The primary objectives of this project include developing an active-learning-based solution to efficiently create large-scale training datasets from PlanetScope time series for specific classes such as burned areas and tree cover. The project also aims to generate high-quality 3-meter resolution training datasets for these classes, provide clean and quality-controlled labeled data to the broader research community, and share the developed algorithms and software through peer-reviewed research publications and open-source code repositories.

To achieve all these objectives, the project has developed an active learning framework based on the U-Net architecture, designed to efficiently generate training data information from PlanetScope imagery. This framework follows a systematic approach.

1) Initial Training: The U-Net model is first trained using either coarser resolution burned/unburned validation data from a globally distributed, manually-labeled set of 30-meter Landsat images or by manually annotating PlanetScope images to classify tree/no-tree classes in various forested environments.

2) Initial Classification and Quality Assessment: The trained U-Net model is then applied to classify a small set of unlabeled PlanetScope images. The resulting classifications are quality-assessed and manually corrected as necessary.
3) Validation: The U-Net model is subsequently applied to classify a predetermined set of validation images, which have been independently annotated, to assess the classification accuracy.
4) Iteration or Completion: If the classification accuracy meets the required standards, the process stops. If not, the corrected classified images from the second step are added to the existing training dataset, and the U-Net model is retrained. This iterative process continues until the desired accuracy is achieved.

The project has made significant progress and achieved several major accomplishments. Notably, it has generated burned area training data for all of Africa, utilizing 575 pairs of two-date PlanetScope images. Additionally, the methodology for generating tree cover training data has been refined through the active learning approach.

The results and methodologies have been disseminated through the publication of two peer-reviewed journal papers, contributing valuable resources and knowledge to the field of land cover classification and disturbance mapping.

5.7. Case study 7: The pangeo-ML project

The Pangeo-ML project builds upon on the foundation of the Pangeo Project to further enhance machine learning (ML) workflows for researchers and data scientists working with many types of complex multi-dimensional datasets. Recognizing the unique challenges in geoscientific ML workflows, such as data dimensionality, transformations, and large volumes, the Pangeo-ML team has focused on developing high-level tools that can bridge the gaps between commonly used geoscientific exploratory data analysis software and deep learning (DL) frameworks.

This project aims to simplify data preprocessing, expand software interoperability, and foster an open-source community equipped with the tools and knowledge to work with Earth observation (EO) data in ML applications. A core objective of Pangeo-ML has been to improve the interoperability of the scientific Python ecosystem, making it easier to construct preprocessing pipelines for ML applications. To this end, the team has contributed to the integration of the Holoviz suite of tools (including hvPlot, GeoViews,

Holoviews, Datashader, SpatialPandas) with other key components of the scientific Python ecosystem, such as Zarr, Xarray, and Rioxarray. This integration has greatly simplified the interactive exploration and preprocessing of Earth science and ML datasets. Additionally, the project has enhanced the interoperability between Xarray, Dask, and geospatial libraries like Pytroll Satpy and Pyresample, streamlining common tasks such as geographic resampling in preprocessing pipelines. Another significant achievement of the Pangeo-ML project is the development of new software interfaces between Xarray and machine learning libraries. The Xbatcher library, a notable outcome of this work, simplifies batch data generation from Xarray datasets, supporting direct integration with popular ML frameworks like TensorFlow and PyTorch.

This library facilitates lazy batch generation, parallel loading, caching, and data loaders, making it easier to handle large datasets in deep learning workflows.

Beyond software development, the Pangeo-ML team has actively engaged with the open-source community, providing expanded documentation, tutorials, talks, and workshops to support scalable machine learning workflows. Their efforts have led to the release of new and improved open-source software, such as Xbatcher and Kerchunk, as well as foundational packages like Xarray and Dask. The team has also developed machine learning applications that both motivate and guide tool development, including a biomass mapping workflow using Landsat and ICESat/GLAS data, a hydrometeorological data assimilation project using FluxNet, a climate downscaling application, and ocean surface current estimation from remote sensing observations. The Pangeo-ML project has made significant strides in improving ML workflows for geoscientific research by enhancing software interoperability, developing new tools, and fostering an active open-source community. Their work has simplified the process of working with complex multi-dimensional datasets, enabling more efficient and scalable ML applications in the geosciences.

5.8. Case study 8: The global vegetation structure (GVS) project

The Global Vegetation Structure (GVS) project focuses on developing machine learning models to integrate data from various remote sensing technologies, aiming to study and map global vegetation structure. By leveraging the strengths of different remote sensing instruments, particularly lidar sensors, radar, and optical sensors, the project seeks to overcome the limitations of individual technologies and create comprehensive, high-resolution maps of vegetation structure and its changes over time.

Lidar sensors, whether deployed on satellites or from airborne campaigns, offer direct measurements of the vertical profile of vegetation, but their availability and coverage are limited. In contrast, radar and optical sensors, while providing indirect estimates of vegetation structure, offer excellent global coverage.

The GVS project aims to combine the detailed, yet sparse, data from lidar with the broad coverage of radar and optical sensors using advanced machine learning techniques, enabling the creation of wall-to-wall maps of vegetation structure on a global scale. The project is structured around several key objectives, beginning with the assessment and inter-calibration of lidar data from different instruments, both space-borne and airborne. This involves collecting and validating airborne lidar campaign data and comparing it with space-borne products. Where necessary, inter-calibration techniques are mainly applied to harmonize data from different types of instruments, ensuring that the aggregated dataset is of high quality and suitable for integration with other types of remote sensing data. Preprocessing of dense predictors is another very crucial step in the GVS project.

This involves developing and applying many types of methodologies to preprocess input datasets, such as optical imagery from Landsat and Sentinel-2, and radar data from the Advanced Land Observing Satellite (ALOS) PALSAR, on a global scale.

The preprocessing ensures that these datasets are aggregated to different resolutions and aligned on compatible grids, facilitating their integration with lidar data. Once the data is collected and preprocessed, the project focuses on testing and comparing various machine learning models. By utilizing the diverse input data sources, the team evaluates different models to identify the most effective approach for estimating vegetation structure.

This process involves assessing the drawbacks and limitations of each method, ensuring a comprehensive evaluation that leads to improvements in the models. The goal is to establish a consistent multiscale approach that can be applied globally while also considering the costs associated with each methodology.

A critical aspect of the project is the intercomparison of the derived products with ground truth airborne datasets. By benchmarking the machine learning model outputs against these ground truth datasets, the GVS project can assess the accuracy and reliability of its methods. This benchmarking is essential for validating the models and ensuring that they provide robust and accurate estimates of vegetation structure on a global scale.

Major accomplishments of the GVS project include the development of a methodology to combine data information from two types of different lidar instruments, the Ice, Cloud, and land Elevation Satellite-2 (ICESat-2) and the Global Ecosystem Dynamics Investigation (GEDI) missions. This approach helps to fill out the observation gaps in GEDI data over boreal areas, enhancing the overall coverage and accuracy of vegetation structure mapping. Multiple machine learning (ML) models for estimating vegetation structure from optical and radar imagery have been tested and applied globally, yielding results that are on par with the current state-of-the-art in the field.

The GVS project represents a very comprehensive effort to advance the study of global vegetation structure through the integration of lidar, radar, and optical remote sensing data. By combining all these types of diverse data sources with cutting-edge machine learning (ML) techniques, the project aims to produce accurate, high-resolution maps of vegetation structure, contributing valuable insights for ecological research and environmental management on a global scale.

5.9. Case study 9: Training data for streamflow estimations

The collaborative project between NASA's Goddard Space Flight Center, the Alaska Satellite Facility, the University of Arizona, and the University of Maryland is focused on developing a comprehensive dataset of river width measurements using ESA Sentinel-1 C-Band Synthetic Aperture Radar (SAR) data. This dataset is intended for training machine learning models that estimate river flow rates and for use in related hydrological models. Sentinel-1 SAR data, with its ability to provide high-resolution, all-weather, day-and-night data at a nominal six-day revisit time, is central to this initiative.

The project has combined several key objectives. First, it aims to provide the research community with a more robust dataset of the river width measurements through NASA's Physical Oceanography Distributed Active Archive Center (PO.DAAC).

This dataset will also be very instrumental in enhancing the overall capability of the research community to map surface water at a 10-meter resolution using Sentinel-1 data distributed by the Alaska Satellite Facility. Additionally, the project also seeks to demonstrate the utility of these river width measurements in deriving accurate river flow rate estimates, which are also very much crucial for various hydrological studies and applications.

To achieve all these objectives, the project is developing a workflow for generating effective river width measurements, defined as the surface water area divided by the river reach length. This workflow is being executed on the Alaska Satellite Facility's (ASF) OpenScienceLab System and comprises three primary components: preprocessing using standard ASF Sentinel-1 methods, a surface water extent mapping program, and a river width measurement program that utilizes the surface water maps as input. An essential feature of the system is a module that filters out Sentinel-1 scenes that do not include river reaches of interest, specifically those listed in the NASA Surface Water and Ocean Topography (SWOT) Mission SWOT River Database (SWORD). The project is evaluating several surface water mapping algorithms using Sentinel-1 SAR data.

These include HydroSAR, developed by the Alaska Satellite Facility; the Equal Percent Solution, developed by the project's principal investigator; and a machine learning algorithm from the University of Arizona. Additionally, the team is considering algorithms from the NASA Observational Products for End-Users from Remote Sensing Analysis (OPERA) Project and a hybrid algorithm that combines elements of the Equal Percent Solution and the OPERA algorithm.

The accuracy of these water maps is now being assessed using hand-labeled water maps derived from high-resolution commercial data provided by the Planetscope constellation. The algorithm deemed most effective will be integrated into the ASF processing system and made available to the research community through an interface provided by ASF. The river width measurement program, a modification of the RivWidth Cloud Program developed by the University of North Carolina, will be adapted to use the Sentinel-1 water maps and interface with the SWORD database.

This integration will automate the measurement of effective river width for the nodes and reaches in the database. The accuracy of the river width data will be evaluated using hand measurements based on Planetscope data. The project will also assess the utility of these river width measurements for deriving river flow rates in a machine learning model and within the SWOT Mission GeoBAM model.

The project has achieved a large number of significant milestones, including the development of a new algorithm for mapping surface water using Sentinel-1 data, known as "The Equal Percent Solution." This algorithm adjusts a radar backscatter threshold to balance false positives and false negatives, improving the accuracy of water detection. A high-resolution dataset of hand-labeled water maps based on Planetscope data has been created and made available to the research community for training machine learning models and evaluating water maps.

The project also implemented an end-to-end system for measuring river width using Sentinel-1 SAR data, although an issue with releasing the river width measurement code to the public led to its removal, with replacement code currently in development.

This project represents a very significant effort to advance the use of remote sensing data for hydrological modeling and river flow rate estimation. By developing and providing high-quality datasets and tools to the research community, the project aims to enhance the overall understanding and management of global water resources.

To provide a better retrospect on the matter of perspectives and how they perform in real time dynamics **Figures 1–3** offers visualizations for a better understanding.

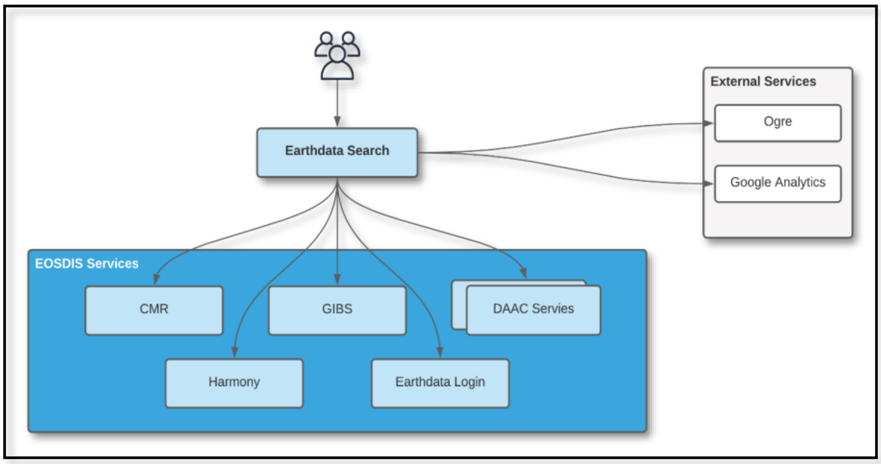

Figure 1. An overview of earth data in action (EOSDIS).

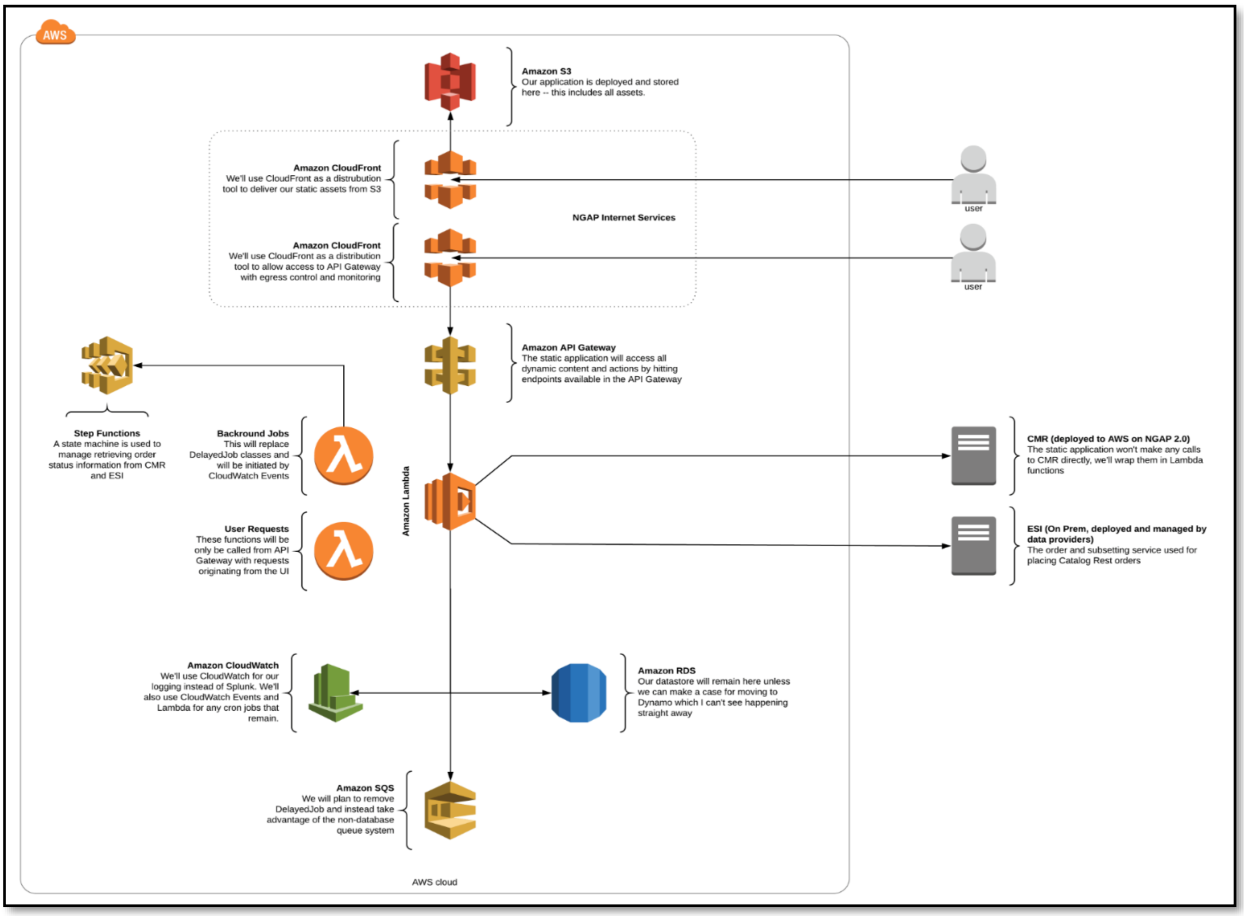

Figure 2. An overview of aws services in action (EOSDIS).

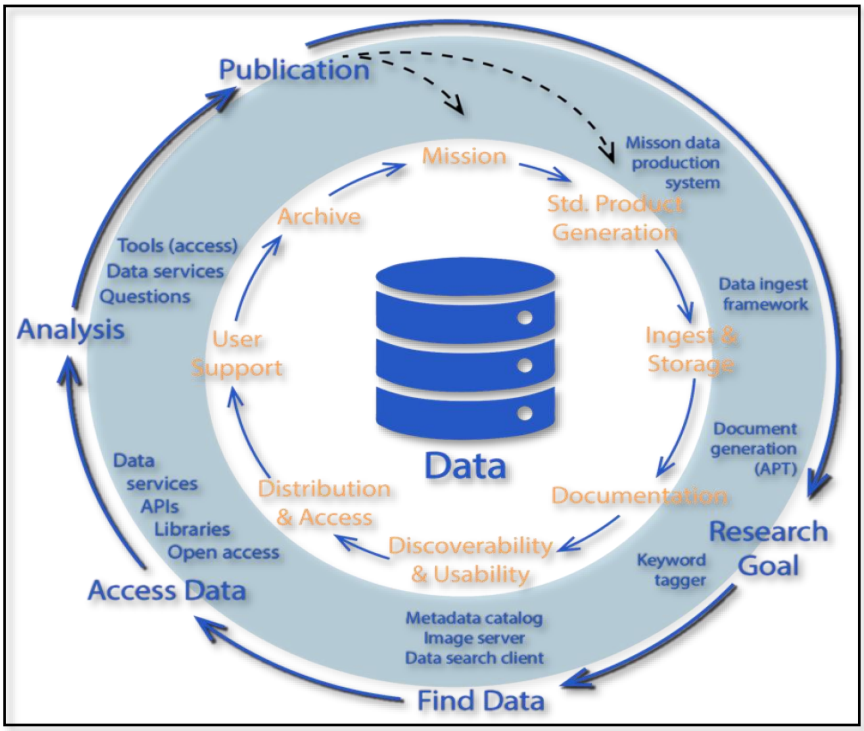

Figure 3. An overview of NASA's AI workflow for earth data.

6. Results and findings

This section presents the detailed results and findings of the research exploration investigations, highlighting the integration of Big Earth Data, machine learning algorithms, and remote sensing technologies for geological and mineral mapping. The outcomes are contextualized to provide a refined clarity on the overall improvements achieved, insights gained, and the various types of associated challenges which were encountered.

6.1. Geological and lithological mapping

The application of machine learning techniques to geological and lithological mapping demonstrated significant advancements in the classification of geological features. For example, using AVIRIS-NG hyperspectral data for mapping gold-bearing granite-greenstone rocks in Hutti, India, support-vector machines (SVM) outperformed other algorithms, achieving an accuracy of 90.3%.

In Brazil's Cinzento Lineament, the combination of spatial constraints with remote sensing data achieved 78.7% accuracy, underscoring the role of integrating spatial data for enhanced results. Similarly, hyperspectral data in Morocco's Central Jebilet region yielded a classification accuracy of 93.05%, slightly higher than the 89.24% achieved with multispectral data. These results validate the robustness of machine learning in processing complex geospatial data.

However, challenges such as vegetation cover significantly impacted the results, necessitating preprocessing techniques like band selection and dimensionality reduction. **Table 1** lists the datasets, models, and performance metrics, providing an overview of the experimental framework and the respective outcomes for each case study.

6.2. Landslide susceptibility and hazard mapping

The integration of topographic and lithological datasets with satellite imagery enabled accurate landslide susceptibility mapping. For instance, in Fruška Gora Mountain, Serbia, the SVM algorithm outperformed Decision Trees and Logistic Regression with an accuracy exceeding 85%, effectively identifying high-risk zones. In Honshu Island, Japan, the combination of ASTER geomorphic data and geological maps with Artificial Neural Networks (ANN) achieved a prediction accuracy of over 90%, showcasing the reliability of machine learning in disaster-prone areas.

These results were validated through the exploration investigations coupled with the associated cross-validation techniques and independent datasets, demonstrating the overall robustness across many types of diverse terrains. The findings emphasize the role of advanced models in urban planning and disaster management.

6.3. Discontinuity analyses

Machine learning, particularly Convolutional Neural Networks (CNNs), excelled in recognizing geological discontinuities like fault planes and bedding planes. For instance, experiments in Korea revealed that CNN-based models achieved a specificity and negative predictive value (NPV) exceeding 0.99, ensuring highly accurate fracture detection even under challenging conditions, such as overlapping geological features or dense vegetation cover. Data augmentation techniques, including flipping and cropping, were instrumental in enhancing model generalization.

6.4. Carbon dioxide leakage detection

Hyperspectral imaging combined with machine learning algorithms identified vegetation stress signals indicative of CO_2 leakage from underground sequestration sites. The ISODATA clustering technique was particularly effective, clustering pixels with similar stress responses to detect leakage zones. In the ZERT site in the US, this method yielded promising results, albeit influenced by seasonal and vegetative variations. These findings highlight the potential for machine learning in environmental monitoring and mitigation efforts.

6.5. Quantification of water inflow in rock tunnels

Using CNNs to classify tunnel face conditions into non-damage, wet, and dripping states achieved an accuracy of 93.01%, significantly improving the automation of water inflow quantification processes. This capability reduces reliance on subjective visual assessments and provides a scalable solution for large-scale infrastructure projects.

6.6. Soil and geological structure classification

Machine learning models demonstrated exceptional performance in classifying soil types and geological structures. For soil classification using Cone Penetration Testing (CPT) logs, ANN models achieved the highest accuracy across various soil types. Similarly, CNNs and Transfer Learning approaches were highly effective in identifying geological structures such as folds, faults, and dikes, achieving classification accuracies of 80%–90%.

7. Earthquake early warning systems and forecasting

Machine learning enhanced earthquake detection and forecasting by effectively distinguishing earthquake signals from noise. Models like Random Forest and GANs accurately recognized P-waves, with laboratory experiments showcasing their ability to predict fault failure time, contributing to improved early warning systems.

The results collectively underscore the transformative role of machine learning in geological and mineral mapping. The AI models demonstrated superior accuracy, efficiency, and adaptability compared to traditional methods. By integrating Big Earth Data, the research addressed challenges such as vegetation cover and data heterogeneity, paving the way for more precise and automated mapping solutions.

The findings revealed that algorithm selection significantly impacts outcomes, with SVMs and CNNs often outperforming other methods for specific applications. However, challenges such as algorithm transparency (e.g., neural networks as "black-box" models) and computational costs require further exploration to optimize their deployment.

Future experimental outlook

To build upon the current findings, future experiments will focus on:

- Enhanced algorithm transparency: Developing interpretable machine learning models to improve stakeholder trust and usability.
- Dynamic data integration: Incorporating temporal changes in Big Earth Data to study evolving geological and environmental processes.
- Scalable solutions: Exploring distributed computing and cloud-based frameworks for real-time data processing and analysis.
- Interdisciplinary approaches: Combining AI with domain-specific knowledge to address complex earth science challenges.

Figures 4–8 visually illustrate key results and findings, while **Table 1** provides a consolidated view of the datasets, methodologies, and performance metrics, facilitating a comprehensive understanding of the research outcomes.

Table 1. The various data sources results and findings in action.

Objective	**Input dataset**	**Location**	**Machine learning algorithms (MLAs)**	**Performance**
Lithological Mapping of Gold-bearing granite-greenstone rocks [46]	AVIRIS-NG hyperspectral data	Hutti, India	Linear Discriminant Analysis (LDA), Random Forest, Support Vector Machine (SVM)	Support Vector Machine (SVM) outperforms the other Machine Learning Algorithms (MLAs)
Lithological Mapping in the Tropical Rainforest [45]	Magnetic Vector Inversion, Ternary RGB map, Shuttle Radar Topography Mission (SRTM), False color (RGB) of Landsat 8 combining bands 4, 3 and 2	Cinzento Lineament, Brazil	Random Forest	Two predictive maps were generated: (1) Map generated with remote sensing data only has a 52.7% accuracy when compared to the geological map, but several new possible lithological units are identified

Table 1. (*Continued*).

Objective	Input dataset	Location	Machine learning algorithms (MLAs)	Performance
				(2) Map generated with remote sensing data and spatial constraints has a 78.7% accuracy but no new possible lithological units are identified
Geological Mapping for mineral exploration [47]	Airborne polarimetric Terrain Observation with Progressive Scans SAR (TopSAR), geophysical data	Western Tasmania	Random Forest	Low reliability of TopSAR for geological mapping, but accurate with geophysical data.
Geological and Mineralogical mapping	Multispectral and hyperspectral satellite data	Central Jebilet, Morocco	Support Vector Machine (SVM)	The accuracy of using hyperspectral data for classifying is slightly higher than that using multispectral data, obtaining 93.05% and 89.24% respectively, showing that machine learning is a reliable tool for mineral exploration.
Integrating Multigeophysical Data into a Cluster Map [48]	Airborne magnetic, frequency electromagnetic, radiometric measurements, ground gravity measurements	Trøndelag, Mid-Norway	Random Forest	The cluster map produced has a satisfactory relationship with the existing geological map but with minor misfits.
High-Resolution Geological Mapping with Unmanned Aerial Vehicle (UAV) [42]	Ultra-resolution RGB images	Taili waterfront, Liaoning Province, China	Simple Linear Iterative Clustering-Convolutional Neural Network (SLIC-CNN)	The result is satisfactory in mapping major geological units but showed poor performance in mapping pegmatites, fine-grained rocks and dykes. UAVs were unable to collect rock information where the rocks were not exposed.
Surficial Geology Mapping [49] Remote Predictive Mapping (RPM)	Aerial Photos, Landsat Reflectance, High-Resolution Digital Elevation Data	South Rae Geological Region, Northwest Territories, Canada	Convolutional Neural Networks (CNN), Random Forest	The resulting accuracy of CNN was 76% in the locally trained area, while 68% for an independent test area. The CNN achieved a slightly higher accuracy of 4% than the Random Forest.
Landslide Susceptibility Assessment [50]	Digital Elevation Model (DEM), Geological Map, 30m Landsat Imagery	Fruška Gora Mountain, Serbia	Support Vector Machine (SVM), Decision Trees, Logistic Regression	Support Vector Machine (SVM) outperforms the others
Landslide Susceptibility Mapping [51]	ASTER satellite-based geomorphic data, geological maps	Honshu Island, Japan	Artificial Neural Network (ANN)	Accuracy greater than 90% for determining the probability of landslide.
Landslide Susceptibility Zonation through ratings [52]	Spatial data layers with slope, aspect, relative relief, lithology, structural features, land use, land cover, drainage density	Parts of Chamoli and Rudraprayag districts of the State of Uttarakhand, India	Artificial Neural Network (ANN)	The AUC of this approach reaches 0.88. This approach generated an accurate assessment of landslide risks.
Regional Landslide Hazard Analysis [53]	Topographic slope, topographic aspect, topographic curvature, distance from drainage, lithology, distance from lineament, land cover from TM satellite images, Vegetation index (NDVI), precipitation data	The eastern part of Selangor state, Malaysia	Artificial Neural Network (ANN)	The approach achieved 82.92% accuracy of prediction.

Table 1. (*Continued*).

Objective	Input dataset	Location	Machine learning algorithms (MLAs)	Performance
Recognition of Rock Fractures [54]	Rock images collected in field survey	Gwanak Mountain and Bukhan Mountain, Seoul, Korea and Jeongseon-gun, Gangwon-do, Korea	Convolutional Neural Network (CNN)	The approach was able to recognize the rock fractures accurately in most cases. The Negative Prediction Value (NPV) and the Specificity are over 0.99.
Detection of CO2 leak from a geologic sequestration site [55]	Aerial hyperspectral imagery	The Zero Emissions Research and Technology (ZERT), US	Iterative Self-Organizing Data Analysis Technique (ISODATA) method	The approach was able to detect areas with CO_2 leaks however other factors like the growing seasons of the vegetation also interfere with the results.
Quantification of water inflow in rock tunnel faces [56]	Images of water inflow		Convolutional Neural Network (CNN)	The approach achieved an average accuracy of 93.01%.
Soil classification [57]	Cone Penetration Test (CPT) logs		Decision Trees, Artificial Neural Network (ANN), Support Vector Machine	The Artificial Neural Network (ANN) outperformed the others in classifying humous clay and peat, while the Decision Trees outperformed the others in classifying clayey peat. Support Vector Machine gave the poorest performance among the three.
Geological structures classification [58]	Images of geological structures		K nearest neighbors (KNN), Artificial Neural Network (ANN), Extreme Gradient Boosting (XGBoost), Three-layer Convolutional Neural Network (CNN), Transfer Learning	Three-layer Convolutional Neural Network (CNN) and Transfer Learning reached accuracies up to about 80% and 90% respectively, while others were relatively low, ranges from about 10% to 30%.
Discriminating earthquake waveforms [59]	Earthquake dataset	Southern California and Japan	Generative Adversarial Network (GAN), Random Forest	The approach can recognise P waves with 99.2% accuracy and avoid false triggers by noise signals with 98.4% accuracy.
Predicting time remaining for next earthquake [60]	Continuous acoustic time series data		Random Forest	The R^2 value of the prediction reached 0.89, which demonstrated excellent performance.
Streamflow Estimate with data missing [61]	Streamgage data from NWIS-Web	Four diverse watersheds in Idaho and Washington, US	Random Forests	The estimates correlated well to the historical data of the discharges. The accuracy ranges from 0.78 to 0.99.

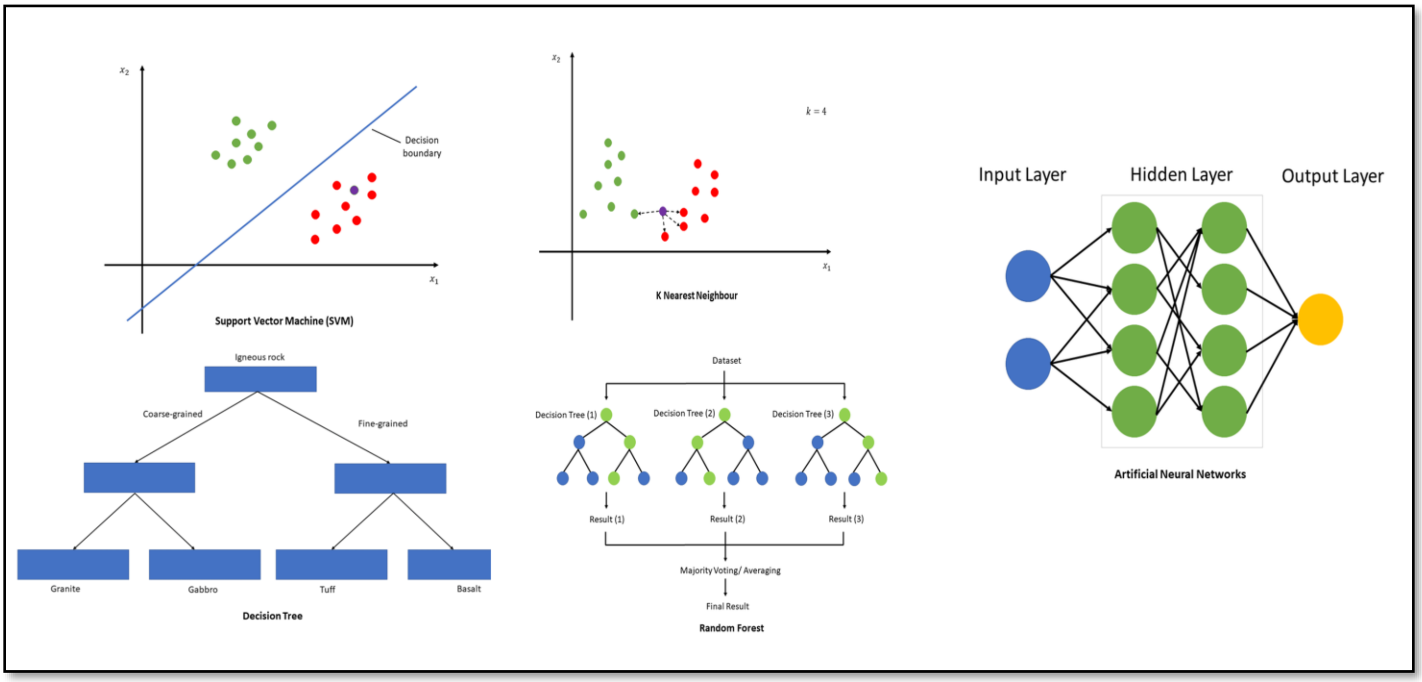

Figure 4. The AI, DL, ML perspectives in action.

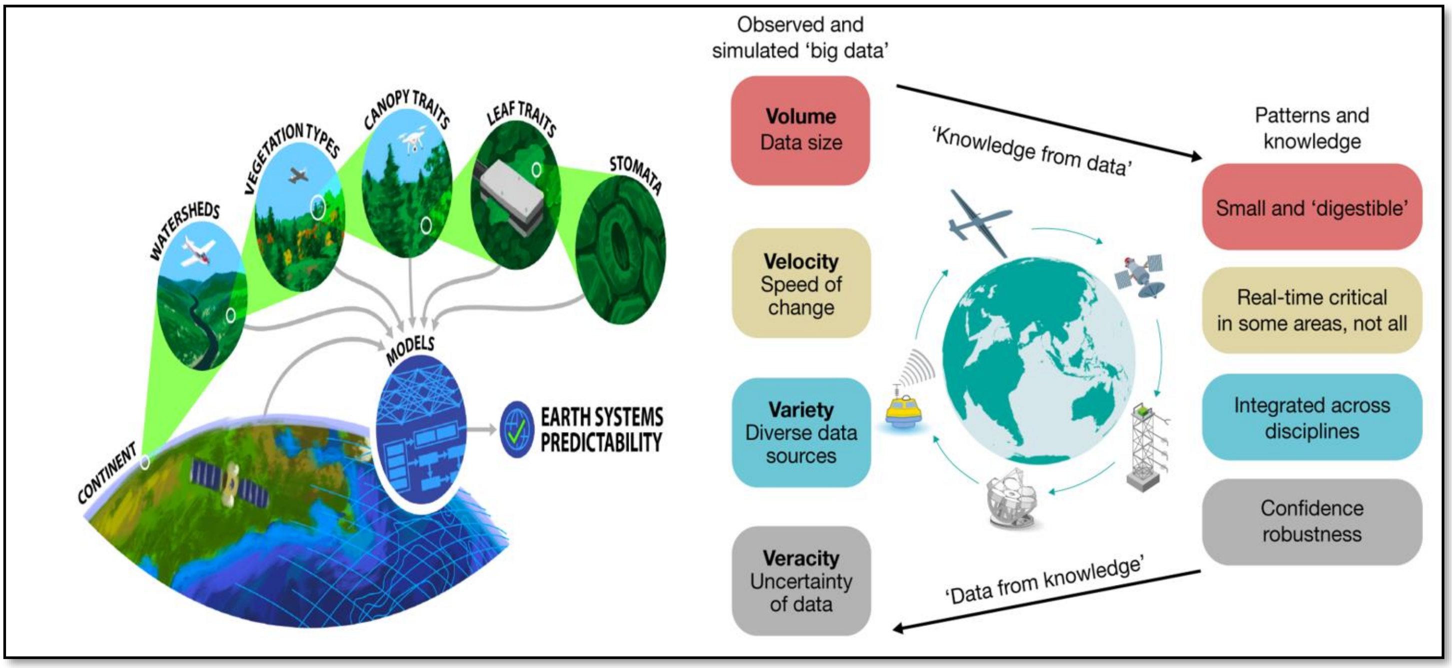

Figure 5. The results and findings from the research explorations 1.

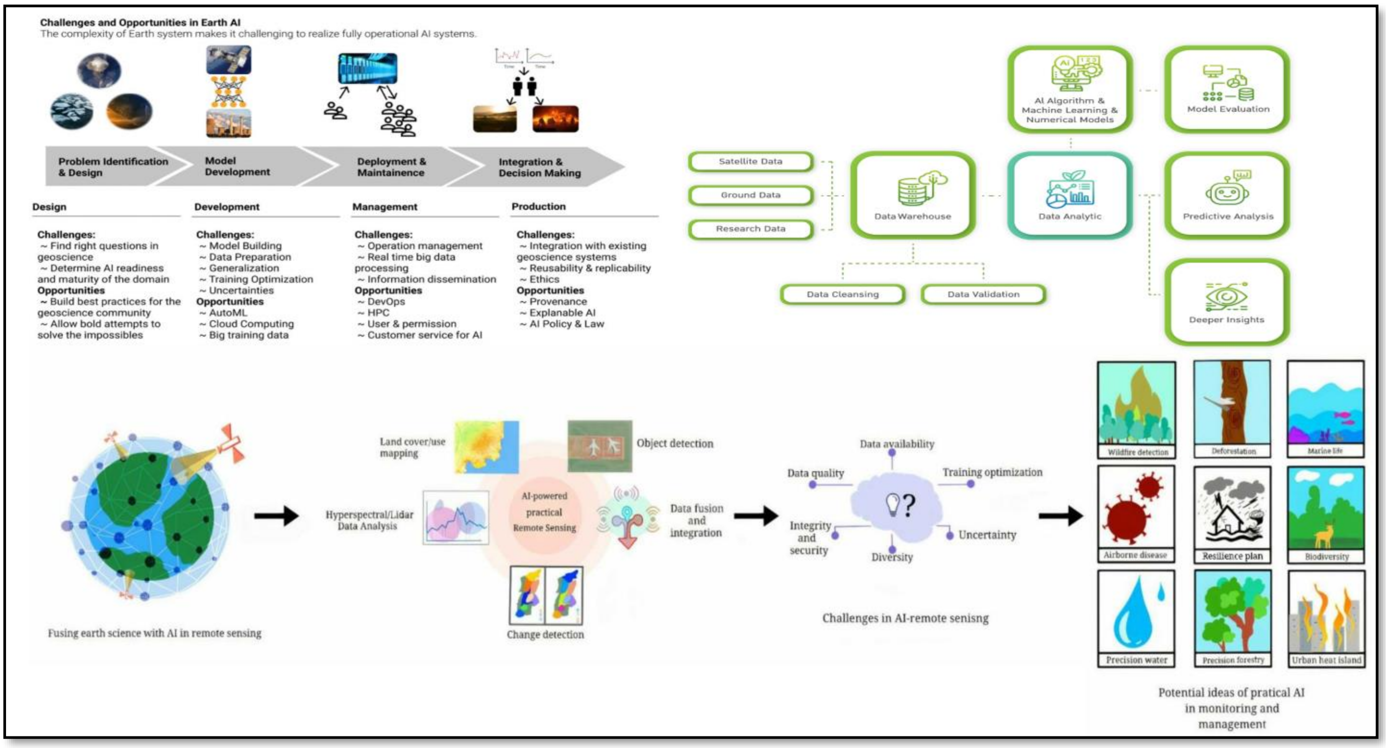

Figure 6. The results and findings from the research explorations 2.

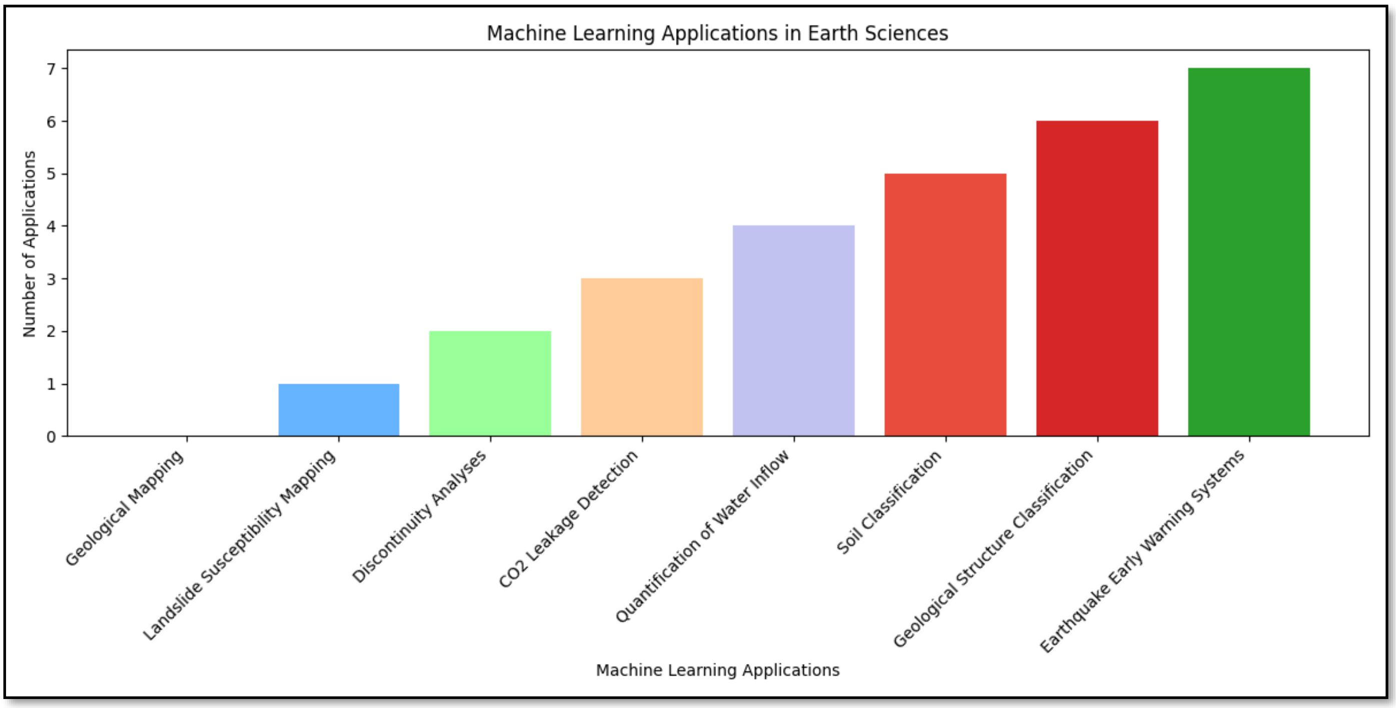

Figure 7. The results and findings from the research explorations 3.

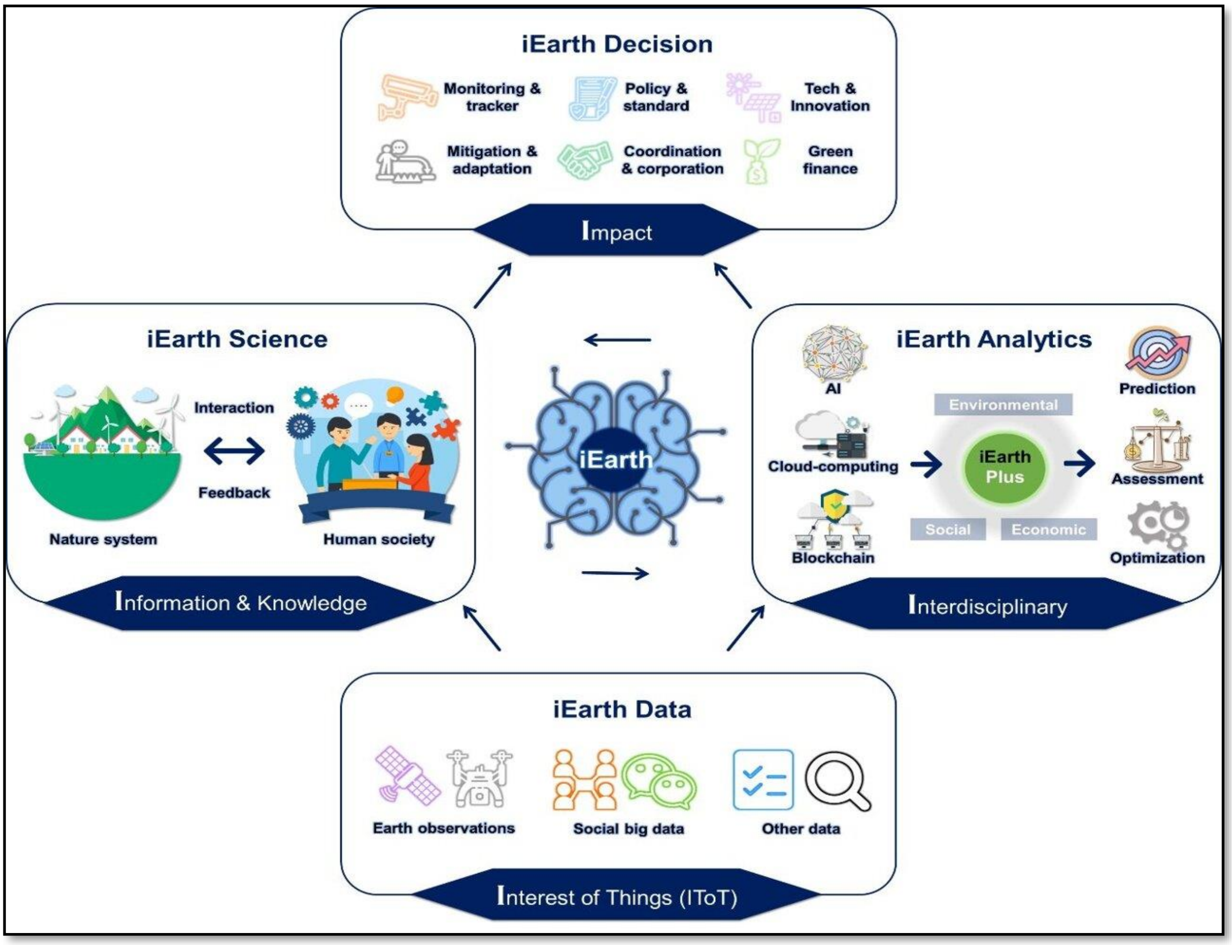

Figure 8. A future outlook for iEarth (experimental).

8. Discussions and future directions

The integration of Big Earth Data and Artificial Intelligence (AI) has revolutionized geological and mineral mapping by addressing the limitations of traditional methodologies, particularly in handling vast and complex datasets. The study highlights the transformative role of machine learning (ML) and deep learning (DL) algorithms in geosciences, offering robust solutions for data-driven analyses.

One of the most notable findings of this research is the demonstrated efficacy of Convolutional Neural Networks (CNNs) in identifying intricate geological formations and mineral deposits. These models excel in capturing subtle spatial patterns, which are often overlooked or misinterpreted during manual analyses. For instance, CNNs successfully analyzed hyperspectral and multispectral datasets [60–66], providing accurate classifications of mineral types and their associated geological features. The fusion of spectral data with AI algorithms has proven particularly effective in regions with dense vegetation cover or complex geological formations, reducing uncertainties and improving prediction reliability. Additionally, the case studies included in this research—spanning diverse geological settings—validate the adaptability of AI-driven approaches. For example, the use of Support Vector Machines (SVM) and Random Forest algorithms demonstrated high accuracy in specific applications, such as lithological mapping and mineral prospectivity.

These findings underline the importance of selecting context-appropriate algorithms to address the unique challenges presented by different geological environments. Despite these advancements, several limitations persist. The variability of Big Earth Data, compounded by noise, artifacts, and limited access to high-quality labeled datasets, poses significant challenges. These issues can lead to overfitting or inaccuracies in AI models. Furthermore, the interpretability of AI algorithms remains a pressing concern. While complex models like neural networks provide superior performance, their “black-box” nature complicates understanding the reasoning behind their classifications, which is critical for decision-making in resource exploration [60–75]. Addressing these challenges necessitates ongoing refinement of both the data inputs and the algorithms themselves. To maximize the potential of Big Earth Data and AI in geological and mineral mapping, the following key areas merit focused exploration and development.

1) Enhanced Data Integration and Fusion

The geosciences field relies on diverse datasets, including seismic surveys, geophysical data, geochemical analyses, and satellite imagery. Future research should prioritize advanced data fusion techniques to integrate these sources seamlessly. By combining multispectral and hyperspectral data with subsurface information, researchers can develop holistic geological models that provide a more accurate representation of subsurface structures and mineral distributions. Techniques such as generative adversarial networks (GANs) and multi-modal learning may offer innovative solutions for handling heterogeneous data sources.

2) Explainable AI (XAI) and Model Interpretability

As AI algorithms grow more sophisticated, there is a critical need for transparency in their decision-making processes. Explainable AI (XAI) tools, such as saliency maps and feature attribution methods, can shed light on the internal workings of AI models. This is especially crucial for applications like mineral prospectivity mapping, where actionable insights are required. Researchers should investigate new frameworks for balancing model complexity with interpretability, ensuring that stakeholders, including geologists and policymakers, can trust and understand the predictions.

3) Scalability and Real-time Processing

The exponential growth of Big Earth Data demands scalable and efficient AI solutions. Future studies should explore distributed computing environments, such as cloud platforms and edge computing, to facilitate real-time data processing. These technologies could enable on-the-fly geological analyses, especially during field surveys or emergency scenarios like landslides. Moreover, developing lightweight AI models optimized for mobile and UAV platforms could revolutionize real-time geological mapping and monitoring.

4) Hybrid AI-Geoscience Approaches

While AI models offer powerful analytical capabilities, geoscientific expertise remains indispensable for contextual interpretation. Future research should focus on hybrid frameworks that combine the strengths of AI with domain-specific knowledge. These approaches could include embedding geoscientific principles into AI algorithms or creating workflows where human expertise complements AI-driven analyses.

Collaborative efforts between AI developers and geoscientists are essential for achieving more reliable and context-aware models.

5) Applications in Sustainable Resource Management

The adoption of AI in geological mapping has significant implications for sustainable resource management. AI can optimize mineral exploration by identifying high-potential areas, thereby minimizing environmental impact and financial costs. Future research should expand on using AI to monitor environmental changes resulting from mining activities, such as soil degradation or water contamination. Additionally, integrating conservation-focused data, such as biodiversity indices, with mineral exploration datasets could promote sustainable practices in natural resource management.

6) Improved Training Data and Validation Methods

The quality of training datasets is paramount for AI model performance. Researchers should focus on generating high-quality labeled datasets through methods such as data augmentation, synthetic data generation, and crowd-sourced labeling. Moreover, robust validation methods, including cross-validation and independent test areas, should be employed to ensure the generalizability of AI models across different geological settings.

7) Ethical Considerations and Policy Integration

As AI technologies become integral to geological mapping, ethical considerations must be addressed. Issues such as data privacy, algorithmic bias, and environmental sustainability should be at the forefront of future research. Policymakers and researchers should work together to establish guidelines for the ethical use of AI in resource exploration and land management.

The integration of Big Earth Data and AI represents a paradigm shift in geological and mineral mapping. While significant challenges remain, ongoing advancements in data integration, model interpretability, and computational efficiency offer a promising trajectory for this interdisciplinary field. By addressing these challenges and exploring future directions, researchers can unlock the full potential of AI-driven geosciences, paving the way for more accurate, efficient, and sustainable resource exploration and management practices.

9. Conclusions

The convergence of Big Earth Data and Artificial Intelligence (AI) has introduced transformative opportunities in geological and mineral mapping, marking a significant milestone in the field of geosciences. This study underscores the pivotal role of AI-driven techniques in addressing the complexities of analyzing large-scale geospatial datasets, enabling enhanced accuracy, efficiency, and depth in geological investigations. The integration of advanced machine learning (ML) and deep learning (DL) methodologies with remote sensing technologies has proven instrumental in uncovering detailed insights into the spatial distribution of geological formations and mineral deposits. A key achievement of this research lies in the successful application of AI models, particularly Convolutional Neural Networks (CNNs), to extract intricate spatial patterns and identify subtle spectral signatures from multispectral and hyperspectral datasets.

These capabilities surpass traditional geological mapping methods, which are often limited by their manual nature and susceptibility to human error. Furthermore, the ability of AI to synthesize diverse data sources and detect relationships within complex geological structures has been a significant advancement, offering precision and reliability in mineral prospectivity analysis.

However, this study also highlights notable challenges that must be addressed to fully realize AI's potential in geosciences. The reliance on high-quality training datasets is critical, as inconsistencies or inadequacies in labeled data can compromise model performance and lead to misclassifications. Additionally, the computational intensity of processing Big Earth Data necessitates scalable solutions, such as distributed computing and cloud-based architectures, to enable efficient analysis. Another pressing concern is the interpretability of AI models, often hindered by their "black-box" nature, which can limit their usability in critical decision-making scenarios.

To overcome these challenges, this research emphasizes the need for continuous innovation in AI techniques and closer collaboration between AI specialists and geoscientists. Enhanced transparency through explainable AI (XAI) frameworks and the integration of domain-specific knowledge into AI workflows are identified as key strategies to improve the interpretability and contextual relevance of AI models. Moreover, advancing data integration and fusion methodologies—incorporating geophysical, geochemical, and remote sensing data—can lead to the development of more comprehensive and holistic geological models.

Looking forward, the potential applications of AI in geosciences are expansive. These include not only refined mineral exploration processes but also contributions to sustainable resource management, real-time geological surveys, and environmental monitoring. The findings of this study highlight the critical importance of balancing technological innovation with ethical considerations and sustainability, ensuring responsible use of AI in resource exploration and management.

By addressing current limitations and harnessing emerging opportunities, researchers and practitioners can unlock the full potential of Big Earth Data and AI in geosciences. This study contributes valuable insights into the academic discourse on AI applications while providing practical recommendations for developing more advanced, reliable, and interpretable geological mapping techniques. These advancements pave the way for a new era of accurate, efficient, and sustainable practices in geological and mineral mapping, reinforcing the transformative role of AI in shaping the future of geosciences.

Author contributions: Conceptualization, ZBA; methodology, ZBA; software, ATR; validation, ZBA; formal analysis, ZBA; investigation, ZBA; resources, ZBA; data curation, ATR; writing—original draft preparation, ZBA; writing—review and editing, ZBA; visualization, ZBA; supervision, ZBA; project administration, ZBA; funding acquisition, ZBA. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: The authors would like to acknowledge and thank the GOOGLE Deep Mind Research with its associated pre-prints access platforms. This research was

deployed and utilized under the platform provided by GOOGLE Deep Mind which is under the support of the GOOGLE Research and the GOOGLE Research Publications under GOOGLE Gemini platform. Also, a huge shout out and thanks to the NASA EARTHDATA open-source database along with their massive NASA EARTHDATA datasets, data models, data tools and NASA EARTH SCIENCE DATA repo for providing access to their enormous pool of scientific data sources. Using their provided platform of datasets and database associated files with digital software layouts consisting of free web access to a large collection of recorded models that are found within research access and its related open-source software distributions which is the implementation and simulation of analytics for the proposed research which was undergone and set in motion. There are many data sources some of which are resourced and retrieved from a wide variety of GOOGLE service domains. All the data sources and various domains from which data has been included and retrieved for this research are identified, mentioned and referenced where appropriate. However, the various types of original data sources some of which are not all publicly available, because they contain various types of private information. The available platform provided data sources that support the research results with findings and information of the research investigations are referenced where appropriate.

Availability of data and materials: The various original data sources some of which are not all publicly available, because they contain various types of private information. The available platform provided data sources that support the findings and information of the research investigations are referenced where appropriate.

Conflict of interest: The authors declare no conflict of interest.

References

1. Paravano A, Patrizi M, Razzano E, et al. The impact of the new space economy on sustainability: an overview. Acta Astronautica. 2024; 222: 162-173. doi: 10.1016/j.actaastro.2024.05.046
2. Earth observation. Available online: joint-research-centre.ec.europa.eu (accessed on 23 November 2022).
3. Newcomers Earth Observation Guide | ESA Business Applications. Available online: business.esa.int (accessed on 23 November 2022).
4. GEO at a Glance. Available online: https://earthobservations.org/ (accessed on 25 August 2020).
5. Pennisi E. Meet the Landsat pioneer who fought to revolutionize Earth observation. Science. 2021; 373 (6561): 1292. doi:10.1126/science.acx9080
6. Eklundh L. Remote sensing and Earth observation. Available online: www.nateko.lu.se (accessed on 27 September 2022).
7. Ashley S. Game changer satellite will measure most of the water on the planet. CNN; 2022.
8. Tomás R, Li Z. Earth Observations for Geohazards: Present and Future Challenges. Remote Sensing. 2017; 9(3): 194. doi:10.3390/rs9030194
9. Balenovic I, Marjanovic H, Vuletic D, et al. Quality assessment of high density digital surface model over different land cover classes. Periodicum Biologorum. 2015; 117(4): 459–470. doi: 10.18054/pb.2015.117.4.3452
10. Appendix A—Glossary and Acronyms. Severn Tidal Tributaries Catchment Flood Management Plan—Scoping Stage. UK: Environment Agency; 2009. https://assets.publishing.service.gov.uk/media/5a7c871ae5274a7b7e321087/Severn_Tidal_Tributaries_Catchment_Management_Plan.pdf
11. Ronald T. Terrain models—A tool for natural hazard Mapping. In: Avalanche Formation, Movement and Effects. IAHS Publ; 1987.
12. Making 3D Terrain Maps, Shaded Relief. Available online: https://www.shadedrelief.com/ (accessed on 11 March 2019).

13. Robert S. Elegant Figures What Not To Do: Vertical Exaggeration. NASA Earth Observatory; 2010.
14. Hargitai H, Willner K, Buchroithner M. Methods in Planetary Topographic Mapping: A Review. In: Planetary Cartography and GIS. Springer International Publishing; 2019. pp. 147–174.
15. Banerdt B. Orbital Laser Altimeter. In: The Martian Chronicle. NASA. 2024; 1(3). https://nssdc.gsfc.nasa.gov/nmc/experiment/display.action?id=1996-062A-03
16. NASA, LOLA. Available online: https://science.nasa.gov/mission/lro/lola/ (accessed on 4 May 2020).
17. Cavanaugh JF, Smith JC, Sun X, et al. The Mercury Laser Altimeter Instrument for the MESSENGER Mission. In: Space Science Review. Springer Nature; 2007.
18. Hargitai H, Willner K, Hare T. Fundamental Frameworks in Planetary Mapping: A Review. In: Planetary Cartography and GIS. Springer International Publishing; 2019. pp. 75–101.
19. ALOS World 3D-30m. Available online: www.eorc.jaxa.jp (accessed on 4 May 2020).
20. ALOS World 3D. Available online: www.aw3d.jp (accessed on 4 May 2020).
21. Emisoft. 8 Things To Look For When Selecting Environmental Data Software. Available online: www.emisoft.com (accessed on 21 April 2020).
22. ESG assets may hit $53 trillion by 2025, a third of global AUM. Available online: https://www.bloomberg.com/professional/insights/trading/esg-assets-may-hit-53-trillion-by-2025-a-third-of-global-aum/ (accessed on 21 April 2020).
23. Kuznetsov VD, Sinelnikov VM, Alpert SN. Yakov Alpert: Sputnik-1 and the first satellite ionospheric experiment. Advances in Space Research. 2015; 55(12): 2833–2839. doi: 10.1016/j.asr.2015.02.033
24. How many Earth observation satellites are orbiting the planet in 2021? Available online: https://geospatialworld.net/blogs/how-many-satellites-are-orbiting-the-earth-in-2021/ (accessed on 18 August 2021).
25. Introduction to satellite. Available online: www.sasmac.cn (16 September 2016).
26. NASA. Drought in Iraq: Image of the Day. Available online: https://archive.org/details/ge_08797 (accessed on 18 August 2021).
27. Earth Observing System Data and Information System. Available online: https://www.nasa.gov/stem-content/earth-observing-system-data-and-information-system/ (accessed on 5 June 2024).
28. Earth Science Data and Information System (ESDIS) Project. Available online: https://www.earthdata.nasa.gov/about/esdis (accessed on 3 May 2019).
29. Behnke J, Kittel D. Looking Back at 25 Years With NASA's EOSDIS Distributed Active Archive Centers. AGU Fall Meeting Abstracts (Report). 2017; 2017: IN43A–0060.
30. Topics | Earthdata. Available online: https://www.earthdata.nasa.gov/ (accessed on 24 August 2021).
31. Wanchoo L, James N, Ramapriyan HK. NASA EOSDIS Data Identifiers: Approach and System. Data Science Journal. 2017; 16.
32. Blumenfeld J. The Global Change Master Directory: Data, Services, and Tools Serving the International Science Community. Earthdata (Report); 2017
33. ECHO—Application Program Interfaces (APIs) for Search and Order Global Change Master Directory. Available online: https://www.earthdata.nasa.gov/news/feature-articles/common-metadata-repository-foundation-nasas-earth-observation-data (accessed on 24 August 2021).
34. Bach D. With Azure, the US Army Corps of Engineers gains a 'powerful tool' for storm modeling. Available online: https://news.microsoft.com/source/features/digital-transformation/with-azure-the-us-army-corps-of-engineers-gains-a-powerful-tool-for-storm-modeling/ (accessed on 5 May 2024).
35. Retooling AI for Earth grants with Microsoft Dynamics 365. Available online: https://www.microsoft.com/insidetrack/blog/retooling-ai-for-earth-grants-with-microsoft-dynamics-365/ (accessed on 5 May 2024).
36. Lunden I. Microsoft launches AI for Earth to give $2M in services to environmental projects. TechCrunch; 2017.
37. Microsoft Planetary Computer. Available online: https://planetarycomputer.microsoft.com/ (accessed on 11 June 2023).
38. Microsoft wants to be a major AI player. In: Here's its master plan. Wired UK. https://www.wired.com/story/inside-microsofts-ai-comeback/
39. Smith B. AI for Earth can be a game-changer for our planet. Available online: https://blogs.microsoft.com/on-the-issues/2017/12/11/ai-for-earth-can-be-a-game-changer-for-our-planet/ (accessed on 11 June 2023).

40. AI for Earth - Microsoft AI. Available online: www.microsoft.com (accessed on 11 June 2023).
41. Mueller JP, Massaron L. Machine learning for dummies. John Wiley & Sons; 2021.
42. Sang X, Xue L, Ran X, et al. Intelligent High-Resolution Geological Mapping Based on SLIC-CNN. ISPRS International Journal of Geo-Information. 2020; 9(2): 99. doi: 10.3390/ijgi9020099
43. Si L, Xiong X, Wang Z, et al. A Deep Convolutional Neural Network Model for Intelligent Discrimination between Coal and Rocks in Coal Mining Face. Mathematical Problems in Engineering. 2020; 2020: 1–12. doi: 10.1155/2020/2616510
44. Merembayev T, Yunussov R, Yedilkhan A. Machine Learning Algorithms for Classification Geology Data from Well Logging. 2018 14th International Conference on Electronics Computer and Computation (ICECCO). Published online November 2018: 206–212. doi: 10.1109/icecco.2018.8634775
45. Costa I, Tavares F, Oliveira J. Predictive lithological mapping through machine learning methods: a case study in the Cinzento Lineament, Carajás Province, Brazil. Journal of the Geological Survey of Brazil. 2019; 2(1): 26–36. doi: 10.29396/jgsb.2019.v2.n1.3
46. Kumar C, Chatterjee S, Oommen T, et al. Automated lithological mapping by integrating spectral enhancement techniques and machine learning algorithms using AVIRIS-NG hyperspectral data in Gold-bearing granite-greenstone rocks in Hutti, India. International Journal of Applied Earth Observation and Geoinformation. 2020; 86: 102006. doi: 10.1016/j.jag.2019.102006
47. Radford DDG, Cracknell MJ, Roach MJ, et al. Geological Mapping in Western Tasmania Using Radar and Random Forests. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2018; 11(9): 3075–3087. doi: 10.1109/jstars.2018.2855207
48. Wang Y, Ksienzyk AK, Liu M, et al. Multigeophysical data integration using cluster analysis: assisting geological mapping in Trøndelag, Mid-Norway. Geophysical Journal International. 2020; 225(2): 1142–1157. doi: 10.1093/gji/ggaa571
49. Latifovic R, Pouliot D, Campbell J. Assessment of Convolution Neural Networks for Surficial Geology Mapping in the South Rae Geological Region, Northwest Territories, Canada. Remote Sensing. 2018; 10 (2): 307. doi:10.3390/rs10020307
50. Marjanović M, Kovačević M, Bajat B, et al. Landslide susceptibility assessment using SVM machine learning algorithm. Engineering Geology. 2011; 123(3): 225–234. doi: 10.1016/j.enggeo.2011.09.006
51. Kawabata D, Bandibas J. Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). Geomorphology. 2009; 113(1–2): 97–109. doi: 10.1016/j.geomorph.2009.06.006
52. Chauhan S, Sharma M, Arora MK, et al. Landslide Susceptibility Zonation through ratings derived from Artificial Neural Network. International Journal of Applied Earth Observation and Geoinformation. 2010; 12(5): 340–350. doi: 10.1016/j.jag.2010.04.006
53. Biswajeet P, Saro L. Utilization of Optical Remote Sensing Data and GIS Tools for Regional Landslide Hazard Analysis Using an Artificial Neural Network Model. Earth Science Frontiers. 2017; 14 (6): 143–151. doi:10.1016/s1872-5791(08)60008-1
54. International society for rock mechanics commission on standardization of laboratory and field tests. International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts. 1978; 15(6): 319–368. doi:10.1016/0148-9062(78)91472-9. https://www.sciencedirect.com/science/article/abs/pii/0148906278914729
55. Bellante GJ, Powell SL, Lawrence RL, et al. Aerial detection of a simulated CO2 leak from a geologic sequestration site using hyperspectral imagery. International Journal of Greenhouse Gas Control. 2013; 13: 124–137. doi: 10.1016/j.ijggc.2012.11.034
56. Chen J, Zhou M, Zhang D, et al. Quantification of water inflow in rock tunnel faces via convolutional neural network approach. Automation in Construction. 2021; 123: 103526. doi: 10.1016/j.autcon.2020.103526
57. Bhattacharya B, Solomatine DP. Machine learning in soil classification. Neural Networks. 2006; 19(2): 186–195. doi: 10.1016/j.neunet.2006.01.005
58. Zhang Y, Wang G, Li M, et al. Automated Classification Analysis of Geological Structures Based on Images Data and Deep Learning Model. Applied Sciences. 2018; 8(12): 2493. doi: 10.3390/app8122493
59. Li Z, Meier M, Hauksson E, et al. Machine Learning Seismic Wave Discrimination: Application to Earthquake Early Warning. Geophysical Research Letters. 2018; 45(10): 4773–4779. doi: 10.1029/2018gl077870

60. Rouet‐Leduc B, Hulbert C, Lubbers N, et al. Machine Learning Predicts Laboratory Earthquakes. Geophysical Research Letters. 2017; 44(18): 9276–9282. doi: 10.1002/2017gl074677
61. Petty TR, Dhingra P. Streamflow Hydrology Estimate Using Machine Learning (SHEM). JAWRA Journal of the American Water Resources Association. 2017; 54(1): 55–68. doi: 10.1111/1752-1688.12555
62. Merghadi A, Yunus AP, Dou J, et al. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews. 2020; 207: 103225. doi: 10.1016/j.earscirev.2020.103225
63. Akhtar ZB, Tajbiul Rawol A. Unlocking the Future for the New Data Paradigm of DNA Data Storage : An Investigative Analysis of Advancements, Challenges, Future Directions. Journal of Information Sciences. doi: 10.34874/IMIST.PRSM/JIS-V23I1.47102
64. Bin Akhtar Z. Artificial intelligence (AI) within manufacturing: An investigative exploration for opportunities, challenges, future directions. Metaverse. 2024; 5(2): 2731. doi: 10.54517/m.v5i2.2731
65. Bin Akhtar Z. From bard to Gemini: An investigative exploration journey through Google's evolution in conversational AI and generative AI. Computing and Artificial Intelligence. 2024; 2(1): 1378. doi: 10.59400/cai.v2i1.1378
66. Akhtar ZB. Unveiling the evolution of generative AI (GAI): a comprehensive and investigative analysis toward LLM models (2021–2024) and beyond. Journal of Electrical Systems and Information Technology. 2024; 11(1). doi: 10.1186/s43067-024-00145-1
67. Unraveling the Promise of Computing DNA Data Storage: An Investigative Analysis of Advancements, Challenges, Future Directions. Journal of Advances in Artificial Intelligence. 2024; 2(1). doi: 10.18178/jaai.2024.2.1.122-137
68. Akhtar ZB. The design approach of an artificial intelligent (AI) medical system based on electronical health records (EHR) and priority segmentations. The Journal of Engineering. 2024; 2024(4). doi: 10.1049/tje2.12381
69. Akhtar ZB. Securing Operating Systems (OS): A Comprehensive Approach to Security with Best Practices and Techniques. International Journal of Advanced Network, Monitoring and Controls. 2024; 9(1): 100–111. doi: 10.2478/ijanmc-2024-0010
70. Akhtar ZB. Generative artificial intelligence (GAI): From large language models (LLMs) to multimodal applications towards fine tuning of models, implications, investigations. Computing and Artificial Intelligence. 2024; 3(1): 1498. doi: 10.59400/cai.v3i1.1498
71. Akhtar ZB, Rawol AT. Harnessing artificial intelligence (AI) for cybersecurity: Challenges, opportunities, risks, future directions. Computing and Artificial Intelligence. 2024; 2(2): 1485. doi: 10.59400/cai.v2i2.1485
72. Akhtar ZB, Tajbiul Rawol A. Enhancing Cybersecurity through AI-Powered Security Mechanisms. IT Journal Research and Development. 2024; 9(1): 50–67. doi: 10.25299/itjrd.2024.16852
73. Zarif BA, Ahmed TR. Artificial Intelligence (AI), Extended Reality (XR): A Biomedical Engineering (BME) Perspective Investigation Analysis. RPC Journal of Bio-Med and Clinical Research. RPC Publishers. 2024; 1(1).
74. Akhtar ZB. Artificial Intelligence (AI) within the Realm of Cyber Security. Insight. Electr. Electron. Eng. 2024; 1(1): 1–11.
75. Akhtar Z. Trends Tech Sci Res Computer Vision and Beyond: A Deep Dive Exploration and Investigation. IEEE Region 10 Trends Tech Sci Res. 2024; 7(3). doi: 10.19080/TTSR.2024.07.555711