Ferreira et al. Environ Sci Eur (2020) 32:120
https://doi.org/10.1186/s12302-020-00397-4

Environmental Sciences Europe

REVIEW

Open Access

Monitoring sustainable development by means of earth observation data and machine learning: a review

Bruno Ferreira[1,2,3*], Muriel Iten[2,4] and Rui G. Silva[3,4]

Abstract

This paper presents and explores the different Earth Observation approaches and their contribution to the achievement of United Nations Sustainable Development Goals. A review on the Sustainable Development concept and its goals is presented followed by Earth Observation approaches relevant to this field, giving special attention to the contribution of Machine Learning methods and algorithms as well as their potential and capabilities to support the achievement of Sustainable Development Goals. Overall, it is observed that Earth Observation plays a key role in monitoring the Sustainable Development Goals given its cost-effectiveness pertaining to data acquisition on all scales and information richness. Despite the success of Machine Learning upon Earth Observation data analysis, it is observed that performance is heavily dependent on the ability to extract and synthesise characteristics from data. Hence, a deeper and effective analysis of the available data is required to identify the strongest features and, hence, the key factors pertaining to Sustainable Development. Overall, this research provides a deeper understanding on the relation between Sustainable Development, Earth Observation and Machine Learning, and how these can support the Sustainable Development of countries and the means to find their correlations. In pursuing the Sustainable Development Goals, given the relevance and growing amount of data generated through Earth Observation, it is concluded that there is an increased need for new methods and techniques strongly suggesting the use of new Machine Learning techniques.

Keywords: Sustainable development, Sustainable development goals, Earth observation data, Machine learning

Highlights

- Sustainable Goals and their universality can only be attained through readily available data from affordable sources such as satellite images and similar commonly available sources.
- Earth Observation is an innovative and accurate approach to address the indicators associated with the Sustainable Development Goals.
- There is an increased need for new methods and techniques to process an ever-growing amount of Earth Observation data.
- Machine Learning techniques are crucial in handling Earth Observation data given the enormous quantity of sources and formats.

*Correspondence: baferreira@isq.pt
[1] ISQ-Intelligent and Digital Systems R&Di, Instituto de Soldadura e Qualidade, Rua do Mirante, 258, 4415-491 Grijó, Portugal
Full list of author information is available at the end of the article

Background

The concept of Sustainable Development (SD) has been developed in 1960 when it became evident that environmental problems can be caused by economic and industrial development. In 1972, a first report was published and presented at UN concerning SD. This report, named as the Meadows Report [1], was strongly criticised at that time since it advocated non-growth to the developing countries [2]. Later in 1987, the Brundtland Report (BR) [3] defined the SD concept as development that meets the essential needs of the present without compromising the ability of future generations to meet their own essential needs. In 2000, the Millennium Development Goals (MDGs), established 8 objectives to tackle poverty and hunger, achieve gender equality and improve the health sector [4]. Until 2015 the MDGs [5] drove the progress of SD, including improvements in health and education services, reduced hunger and equity gaps, and higher levels of coverage in interventions with major investments [6, 7]. However, it remained incomplete and in 2012, new objectives were established, designated as Sustainable Development Goals (SDGs) [5], defining 17 unique objectives, representing an urgent call to shift the world onto a more sustainable path [8, 9].

Earth Observation (EO) plays a major role in supporting progress towards many of the SDGs [10, 11]. According to the United Nations [12] it is advantageous using EO data such as the images from satellites to produce and support official statistics to complement traditional sources of socio-economic and environmental data. Satellite imagery may be perhaps the only cost-effective technology able to provide data at a global scale [13, 14]. Such globally available data are determinant to understand the progress and contribution of underdeveloped countries concerning SD since they lack the resources to collect relevant information. The considerable amount of data, provided by EO sources, need to be effectively analysed and processed with appropriate methods and tools to provide robust indicators concerning SD.

The growth of Machine Learning (ML) field, which is constantly creating new opportunities for monitoring and evaluating humanitarian efforts, plays an essential part in the analysis of satellite images applied to SDGs. In fact, the majority of methods used for processing EO data are based on ML [11, 15] given in one hand their ability to process enormous amounts of data and also because they possess unique characteristics pertaining to classification, modelling and forecasting.

The main purpose of this article is to explore and comprehend the relation between SD, EO and ML, to understand the relevance and role EO and ML play in attaining the SDGs. Figure 1 depicts the layout of this review as well as major aspects pertaining to the treatment of EO data related to the identification of SDGs.

This review highlights major methodologies and ML methods that have been successfully applied to EO data in pursue of SD. The structure of this paper is divided as follows: Sect. "Materials and methods" describes how the research was conducted, Sect. "Overview on sustainable development" presents the meaning of SD, its history, concepts and goals, followed by a brief explanation of the EO system and how it presently contributes to SDGs, in Sect. "Overview on earth observation for sustainable goals development". Afterwards, in Sect. "Earth observation using machine learning techniques", a review on the importance of ML for EO is presented and as well as their contribution for SDGs, highlighted by case studies of different ML categories applied to EO data. In addition, further considerations are addressed and discussed concerning SD, EO, ML, their relation and new paths and approaches to overcome limitations.

Materials and methods

A systematic search and analysis of published articles in peer-reviewed journals have been conducted using ScienceDirect and Google Scholar. The search has been performed using the following search topics: *sustainable development* or *sustainable goals*, *earth observation* and *machine learning*. To ensure the identification of

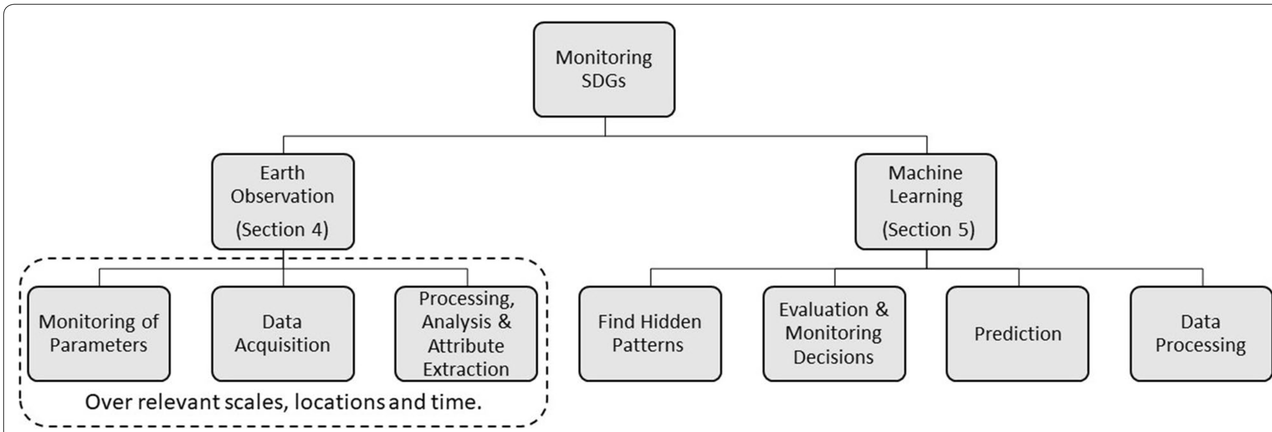

Fig. 1 Relation between SDG, EO and ML

relevant case studies for each ML category of data and image analysis, words such as: *classification techniques*, *clustering techniques*, *regression techniques*, *dimension reduction techniques*, *empirical and semi-empirical modelling*, *supervised techniques*, *unsupervised techniques* and *object-based techniques* in combination with *earth observation data* or *sustainable development goals* were used. The search was refined to sustain relevance and state of the art results, considering the latest research and case studies, retaining historical reports and agreements.

Overview on sustainable development

The environmental problems derived from the economic development became evident during the 1960s and a number of solutions were proposed [1, 3, 8]. The Limits to Growth, also known as the Meadows Report [1], was published by the Club of Rome, in 1972. It presented a computer model developed by MIT called World3, which allowed Meadows et al. [1] to explore the relationship between five subsystems of the world economy: population, food and industrial production, pollution and consumption of non-renewable natural resources [16]. The key finding has been that unlimited growth in the economy and population would lead to a collapse of the global system by the mid to late twenty-first century [1, 17–19]. Moreover, the sooner the world starts striving to change the growth trends, the better the chance of achieving sustainable ecological and economic stability [1, 18, 19]. Thus, the report advocated that the non-growth in developing countries is a response to environmental decline and the lack of its resources [1, 20].

This premise became very popular among non-orthodox economists since it was translated as an attack to the capitalist economic system. On the other hand, it has also been criticised by the economists who affirm that for capitalism, it is crucial a development without boundaries. Due to that, in 1974, the Club of Rome issued another report in which it defended an organic growth (world division into different regions, each with a definite function within the world system) [2]. Since the publication of The Limits to Growth [1], a considerable number of concepts have been introduced and developed integrating ecological and economics concerns, not being consensual, until the publication of Brundtland Report (BR) in 1987 (further detailed in Sect. "Brundtland report [3]"). Table 1 presents some of the most important milestones of the path to SD until nowadays.

Brundtland report [3]

The Brundtland Report's (BR) concept of SD follows a generic definition of development that meets the essential needs of the present without compromising the ability of future generations to meet their own essential needs [3]; however, it included crucial features such as environmental preservation and meeting the basic human needs at a global scale. For those reasons, it was widely accepted as a reference for SD definition [20]. Even so, the ambiguity in the BR's concept of SD along with differing worldviews, ideologies, backgrounds, beliefs and interests has contributed to the proliferation of several explanatory definitions [23]. In an attempt to clarify and simplify the BR's concept, it became important to describe and explain the following key concepts:

- Needs: necessary or basic needs (especially referring to developing countries' needs);
- Technological Limitation: insufficient technological development;
- Social Organisation Problems: originate an unequal allocation of income.

Later, the BR also clarified the meaning of technological growth, arguing that such progress cannot exceed the limited availability of resources [3, 20].

Millennium development goals (MDGs)

In 2000, the Millennium Development Goals (MDGs) have started a global effort to tackle the indignity of the poverty problem. The MDGs [24] established eight objectives for: tackling poverty and hunger; primary education for all children; achieve gender equality; improve maternal and child health; prevent and combat deadly diseases; ensure environmental sustainability; and, global development.

Until 2015, the MDGs allowed progress in several important areas, such as: reducing poverty and child mortality; providing access to water and sanitation; improving maternal health and combatting several diseases such as HIV/AIDS, malaria and tuberculosis.

The most notable accomplishments were: the reduction of child mortality and the number of children out of school by more than half; more than 1 billion people left extreme poverty; and, HIV/AIDS infections have been reduced by almost 40%. The legacy and achievements of the MDGs provided valuable lessons and experience, and pave the way for new goals [8].

Sustainable development goals

The Sustainable Development Goals (SDGs) have replaced the MDGs in 2012 during the UN Conference on SD held in Rio de Janeiro. As a result of climate changes and other serious environmental problems, there was a need to enhance the environmental performance [25]. Hence, the main objective was to create new goals that would address the urgent environmental, political and economic challenges affecting the world [26].

Table 1 Important milestones on the path to SD adapted from: Klarin [21]

Year	Activity	Description
1969	UN published the report Man and His Environment or U Thant Report [22]	Activities focused on avoiding global environmental degradation. More than 2000 scientists were involved in the development of this report
1972	First UN and UNEP world Conference on the Human–Environment, Stockholm, Sweden	Under the slogan Only One Earth, a declaration and action plan for environmental conservation are presented
1975	UNESCO conference on education about the environment, Belgrade, Yugoslavia	Sets up a global environment educational framework, a statement known as the Belgrade Charter
1975	International Congress of the Human–Environment, Kyoto, Japan	Emphasises the problems as in Stockholm in 1972
1979	The First World Climate Conference, Geneva, Switzerland	Focused on the promotion of climate change research and monitoring
1981	The first UN Conference on Least Developed Countries, Paris, France	A report with guidelines and strategies for helping the underdeveloped countries in pursue of SDG
1984	Establishment of UN World Commission on Environment and Development (WCED)	Establishes the cooperation scenario between developed and developing countries and the adoption of global development plans on environmental conservation
1987	WCED report Our Common Future or BR	A report with the foundations of SD's concept
1987	Montreal Protocol	Contains research results on adverse impacts on the ozone layer
1990	The Second World Climate Conference, Geneva, Switzerland	Presents further developments on the climate change research and monitoring, including the creation of a global Climate Change Monitoring System
1992	United Nations Conference on Environment and Development (Earth Summit or Rio Conference), Rio de Janeiro, Brazil	The Rio Declaration and Agenda 21 Action Plan establishes SD principles and a framework for future tasks
1997	Kyoto Climate Change Conference, Kyoto, Japan	The Kyoto Protocol agreement between countries to promote CO_2 reduction and other greenhouse gas emissions, starting in 2005
2000	UN Millennium declaration	Declaration containing 8 MDGs aimed to be accomplished by 2015
2002	The World Summit on SD, Johannesburg, South Africa	Report with the results achieved since the Rio Conference, reaffirming previous obligations and setting the guidelines for future developments
2009	The Third World Climate Conference, Geneva, Switzerland	Further development of the global Climate Change Monitoring System, including early detection of possible disasters.
2009	World Congress Summit G20, Pittsburgh, USA	Agreement amongst G20 member states on a moderate and sustainable economy
2012	UN conference Rio +20, Rio de Janeiro, Brazil	"The Future We Want" reinforced the commitment to the SDGs and encouraged the global green economy
2015	UN SD Summit 2015, New York, SAD	Presents the UN 2030 Agenda for SD setting up 17 SDGs which should be achieved by 2030
2015	COP21 Paris Climate Change Conference, France	Agreement on the reduction of greenhouse gases to mitigate and minimise global warming
2019	COP25 Madrid Climate Change Conference, Spain	Agreement on the reduction of greenhouse gas emissions to zero by 2050—The European Green New Deal

Representing an urgent appeal to change the world's course into a more sustainable direction, the SDGs [27] represent a strong commitment to proceed the MDGs and tackle some of the world's most significant challenges [28].

The success of each of the 17 goals affects all other positively: No Poverty; Zero Hunger; Good Health and Well-Being; Quality Education; Gender equality; Clean Water and Sanitation; Affordable and Clean Energy; Decent Work and Economic Growth; Industry, Innovation and Infrastructure; Reduced Inequalities; Sustainable Cities and Communities; Responsible Consumption and Production; Climate Action; Life Below Water; Life on Land; Peace, Justice and Strong Institutions; Partnerships for the Goals [8]. The 2030 Agenda [5], which coincided with another historical agreement achieved at COP21 Paris Climate Conference [29], sets specific objectives and attainable targets for the reduction of carbon emissions, management of climate change and risks of natural disasters.

Overall, the SDGs are special because they address issues that affect the entire world and reaffirm the determination to eradicate poverty, improve the health system and reduce inequalities. Better yet, they involve all nations in building a more sustainable, safer, more prosperous planet for humanity [8, 28]. To monitor and achieve the SDGs, EO became a vital part since it provides numerous benefits [10, 30, 31], namely: Data at different scales (local, regional, national or even global) and periods of time; Consistency; Wide variety of parameters; and, Cost-effective data acquisition.

Overview on earth observation for sustainable goals development

Earth Observation (EO) covers different approaches, including the use of drones, aircrafts and satellites. The era of satellite based EO began in 1959 with the launch of Explorer 7, and remains until today [32]. In fact, there are more than 2000 active EO satellites operated by Space Agencies, governmental institutions and commercial operators [11, 33], resulting in an increased availability of information concerning the Earth condition and proprieties [34].

EO data are an example of a big data source that can be acquired at low cost, over long periods of time and used to comprehend the entire Earth system while addressing scientific challenges [35] such as climate change and global warming [36], ecological change and reduction impacts of habitat and biodiversity deterioration [37] and used to produce statistics and indicators that enable the quantification of SD [11, 12]. The United Nations report [12] has demonstrated the viability of using EO data to produce official statistics, including SDGs statistics such as agricultural [38], urban and land planning [39] or food security indicators [40].

EO satellite imagery can be classified into two groups, based on the sensor used to capture images: the passive sensors receive emitted or reflected radiation by the Earth's surface, and the active sensors emit radiation and receive the echoes reflected or refracted by the Earth's surface [11]. Overall, EO sensors provide data at four different resolutions: spectral, spatial, radiometric and temporal. The spectral resolution is the ability to define/distinguish wavelengths ranges of radiation; hence, different spectral bands provide a spectral signature for specific land cover types [11] such as soil [41], water [42] or buildings [43]. The spatial resolution refers to the area that each pixel represents on the surface, the radiometric resolution indicates the degree of light intensities the sensor is able to distinguish [44] and the temporal resolution is related to the revisit time, namely the frequency with which sensors cross a specific area on Earth. Besides the differences related to the type of EO sensors, the data provided by satellites can also be distinguished by the different orbits. The geostationary orbit means that satellites track the same area and the Low Earth orbit means that satellites track the surface as they orbit [11].

EO images can be used to identify characteristics of interest based on how images are presented and their inherent properties, such as in agriculture [45], forests [46], water [47] and urban areas [48]. Identifying such characteristics has been often seen as a classification problem which requires techniques to classify or group pixels, according to their spectral characteristics, as belonging to a class [48]. The study of Group on Earth Observations [10] has identified SDGs that are measurable, at some level, using EO data. Figure 2 presents SDGs that can already be measured and analysed based on EO data as SDG 2—No hunger, SDG 6—Clean Water and Sanitation, SDG 13—Climate Action, and SDG 14—Life Below Water.

Taking advantage of emerging developments within EO domain represents an accurate and reliable way to address the SDG indicators and targets and thus bridge the gap between developed and developing countries discrepancy on the quantity and quality of data [20]. The data from EO sources have been advocated by several international organisations and researchers, such as Holloway et al. [38] and Murthy et al. [14], as a mean of minimising costs compared to the conventional acquisition and monitoring of different environmental parameters over relevant scales, areas and time periods [11].

From Fig. 2, it can be depicted that EO can provide quite a large number of indicators for the SDG framework such as data on the condition of the atmosphere [49], oceans [50], crops [51], forests [52], climate [53], natural disasters [54], natural resources [55], urbanisation [56], biodiversity [57] and human conditions [58]. The two most important indicators are population distribution (I-1), and cities/infrastructure mapping (I-2) since they contribute to all the SDGs. On the other hand, the SDGs which benefit from all the EO indicators are the zero hunger (SDG 2), clean water and sanitation (SDG 6), climate action (SDG 13), life below water (SDG 14) and partnership for the goals (SDG 17). This view is supported by the Global Working Group on Big Data [59] and United Nations [12] that states that satellite imagery has significant potential to provide more timely information, minimising the number of surveys and offering more disaggregated data for informed decision making. As a consequence of the quantity of data generated by EO sources, the necessity to find methods to process and analyse this amount of data arises. The purpose is to transform the EO data into valuable information.

Earth observation using machine learning techniques

In the last decade, there have been some major contributions to a wide range of Earth Science applications, from analysing gases, soil, vegetation, climate and, more recently, to ocean [60, 61]. Recent advances on Machine Learning (ML) field are creating unprecedent opportunities to evaluate and monitor policy decisions as well as humanitarian initiatives [62, 63]. Despite the advantages

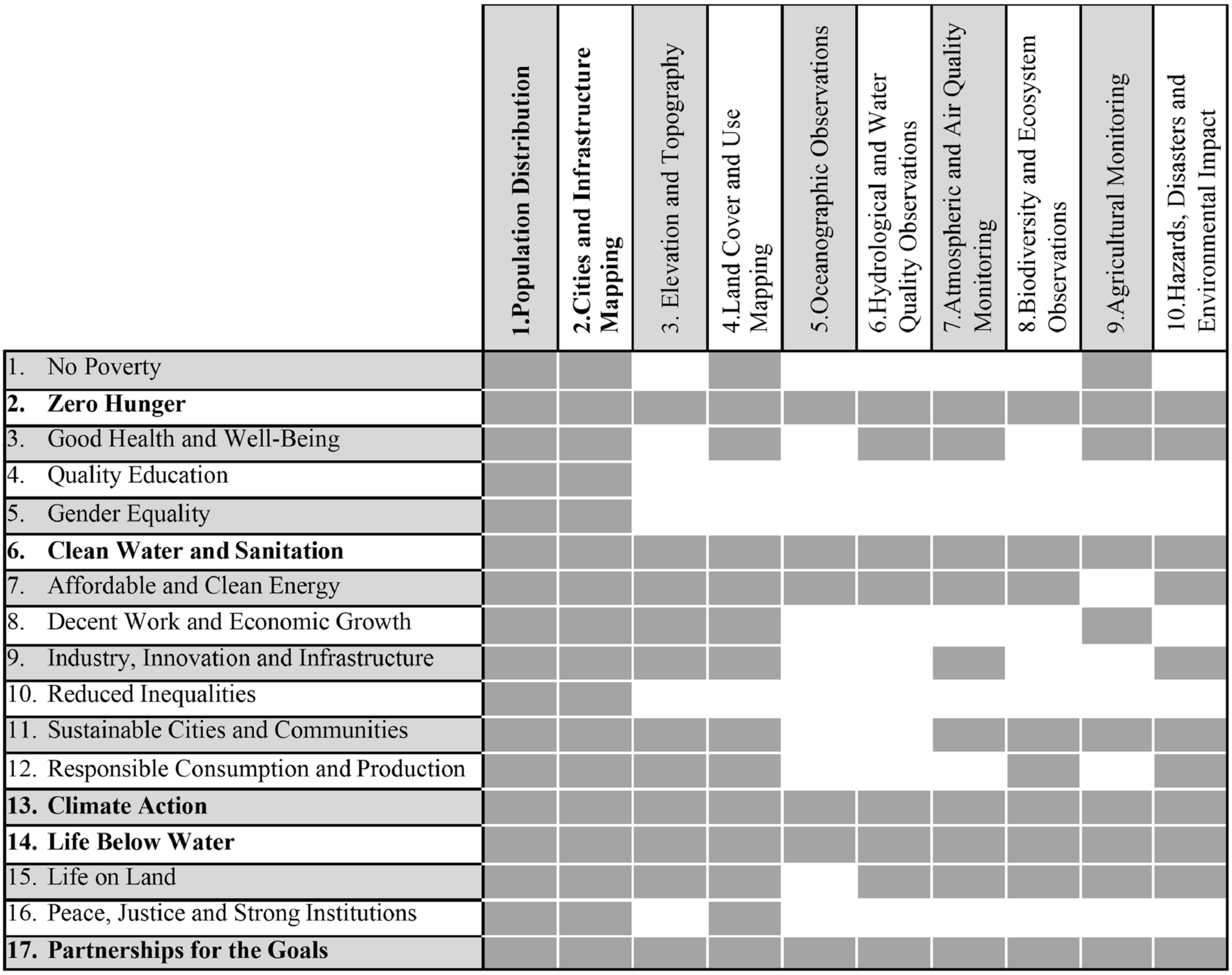

Fig. 2 SDGs measurable by EO data adapted from: Group on Earth Observations [10]

of using ML techniques, it may require greater computational resources as well as an expert to interpret results. ML techniques can be classified into four groups: supervised, unsupervised, semi-supervised and reinforcement learning schemes. The major difference between supervised and unsupervised lies in the fact that the first one requires output values (classification) in the training dataset [64] where problems can be either as classification or regression techniques. In contrast, unsupervised learning techniques require only the input values in the training dataset since their purpose is to find hidden patterns in data and can be handled by clustering or dimension reduction techniques [65]. Semi-supervised learning combines aspects of supervised and unsupervised learning and requires a combination of data with and without classification [66]. Reinforcement learning aims to build systems that can learn from the interaction with the environment, using rewards and punishments rules [67, 68]. The following sub-sections give an overview of the different techniques and methods pertaining to the use of ML in the scope of SD supported in EO data highlighting major findings and applications. This summary outlines the boundaries of research concerning the application of ML algorithms as well as their importance, relevance and potential to support further research towards the development of robust methodologies concerning universal applications. This overview takes into consideration the most recent research results as well as their relevance.

SDGs tackled with machine learning

ML is a subdomain of Artificial Intelligence, which according to Samuel [69] aims to provide to machines the ability to learn from data without being explicitly programmed. The study and development of algorithms plays a major role in ML, as it aims to build a model between inputs and outputs, based on the data and algorithms provided, to learn how to make decisions upon unseen information [70, 71]. The popularity of ML is vast and increasingly applied to different subdomains, including Statistical Learning methods, Data Mining, Image Recognition, Natural Language Processing and Deep Learning [72].

A substantial number of ML algorithms have been used and described in the literature, performing a wide range of tasks in a variety of domains like Agriculture [73], Renewable Energies [74], Disasters [54], Climate [75], Construction [76], Human Living Conditions [58] and Health System [77]. Figure 3 presents the most relevant techniques applied to remote sensed data, grouped according to the four categories of supervised and unsupervised methods: classification, clustering, regression and dimension reduction.

Classification
A classification method belongs to supervised learning category, and it is applicable in cases where the overall aim is to accurately assign a datapoint to a class [78–80]. There is a broad range of classification methods as presented in Table 2, in the scope of SD, that clearly shows the impact and potential use of these techniques in conjunction with EO data.

Clustering
The clustering method belongs to unsupervised learning category, and it is appropriate when the purpose is to associate/divide datapoints into clusters [78, 89]. Table 3 synthesises the findings within the scope of clustering methods used in combination with EO data to aid in the development of SDGs.

Regression
A regression method belongs to the same category as the classification method, supervised learning, and it is applicable when the aim is to predict/estimate a continuous output variable of a given datapoint [78, 99]. There are several approaches, as presented in Table 4, in the scope of SD, that clearly show the impact and potential use of these techniques in conjunction with EO data.

Dimension reduction
Dimension reduction, similar to clustering method, belongs to the unsupervised learning category and typically follow two main approaches: Feature Selection (FS), applicable when there is the necessity to select fewer characteristics [111, 112]; and Feature Extraction, when the information needs to be synthesised through transformation. The aim is to create a small set of features covering much of the details in the initial dataset [79, 113, 114]. Then, these features/characteristics can be fed into other algorithms or otherwise used as an end result [78]. Table 5 synthesises the finding within the scope of dimension reduction methods used in combination with EO data to aid in the development of SDGs.

Methodologies and techniques for EO imagery analysis
Pre-processing, post-processing and the seldom incorporation of qualitative information play a major role in the success of any data analysis approach and is found to vary significantly among researchers. As above mentioned, the

Supervised	Unsupervised
Classification (section 5.1.1)	**Clustering (section 5.1.2)**
▪ Support Vector Machines (SVM) ▪ Classification Trees ▪ Random Forest (RF) ▪ K-Nearest Neighbour (K-NN) ▪ Naive Bayes ▪ Maximum Likelihood Classification ▪ Convolutional Neural Network (CNN)	▪ Gaussian Mixture Models ▪ K-Means ▪ Self-Organising Map ▪ Density-Based Spatial Clustering of Application with Noise (DBSCAN) ▪ Hierarchical Cluster Analysis ▪ Fuzzy C-Means (FCM)
Regression (section 5.1.3)	**Dimension Reduction (section 5.1.4)**
▪ Linear Regression ▪ Regression Trees (RTs) ▪ Support Vector Regression ▪ Neural Networks (NN) ▪ Gaussian Process Regression (GPR) ▪ Kernel Ridge Regression ▪ Partial Least Square Regression	▪ Principal Components Analysis (PCA) ▪ t-Distributed Stochastic Neighbour Embedding (t-SNE) ▪ Discriminant Analysis ▪ Isomap ▪ Autoencoder ▪ Locally Linear Embedding (LLE)

Fig. 3 Categories of ML problems and examples of methods adapted from: Holloway and Mengersen [78]

Table 2 Examples of application of classification methods towards SDGs using EO data

SDGs	Field	Main finding	References
SDG 2 (Zero Hunger)	Agriculture	Multi-temporal crop classification reduces the unfavourable effects of using single-date acquisition	[81]
		The proposed method performed similar to SVM and RF in the classification of crops with similar phenology	[57]
		Developed an efficient framework for multi-temporal crops classification	[82]
SDG 6 (Clean Water and Sanitation)	Wetland	The developed framework for coastal plain wetlands classification had high accuracy.	[83]
SDG 8 (Decent Work and Economic Growth)	Slavery	The approach was used to help to liberate slaves by mapping brick kilns.	[58]
SDG 11 (Sustainable Cities and Communities)	Land use	The approach based on CNN achieved an accuracy of $\cong$ 98% for land use and land cover analysis	[84]
		The proposed approach confirmed its suitability for urban planning because it had a superior performance compared to the global one	[56]
	Living conditions	Deep learning demonstrated a high potential to map areas of deprived living conditions	[85]
	Land cover	The multivariate time series algorithm showed high accuracy for rare land cover classes	[86]
SDG 13 (Climate Action)	Climate	The model based on decision trees, and used to classify local climate zones, achieved a good performance	[75]
SDG 14 (Life Below Water)	Marine habitat	SVM and K-NN classifiers achieved an accuracy higher than 90% on mapping coastal marine habitat	[50]
SDG 15 (Life on Land)	Land cover	The approach used allowed to differentiate the hyperspectral subclasses from the classes	[87]
	Forest	Sentinel-2 is considered a powerful source of data for forest monitoring and mapping	[52]
		RF was the best method to predict and map the area and volume of eucalyptus	[88]

majority of methods for processing EO data are based on ML algorithms, whether they are supervised or unsupervised [11]. However, besides the general problem category, the techniques can also be classified according to the approach used taking into consideration images analysis and their feature extraction: Sub-PB, PB, Super-PB and OB [129, 130]. In Sub-PB, each pixel can have multiple classes [131, 132]; in PB, it is only possible to have one class per pixel [133, 134]; in Super-PB, the pixels are grouped based on homogeneity [135, 136]; while in OB, the aim is to delineate readily usable objects from imagery or partitioning an image into objects [137, 138]. Figure 4 illustrates the Sub-PB, PB, Super-PB and OB techniques.

In addition to those techniques, there are visual interpretation techniques conducted through direct operator analysis of characteristics from raw satellite images. Such techniques are used to extract visual characteristics including colour, form, size, pattern, texture and shadow from images [11]. The human abilities, however, should be explored/emulated to further enhance and automate ML algorithm-based image interpretation. Overall, several approaches are being used by different researchers that combine ML algorithms and pre-processing of data giving rise to different methodologies.

Empirical and semi-empirical modelling

Empirical and Semi-Empirical models are created based on data acquired from observations or experiences, which means that there are none or few assumptions on data analysis. There are many examples of the application of empirical and semi-empirical modelling, such the ones in Table 6:

Supervised classification techniques

The Supervised Classification requires a set of classified samples (sub-pixels, pixels or super-pixels) to train the models to understand each class' patterns. After training models should be able to categorise new samples or place those samples into classes [143]. Some applications of these approaches are presented on the following Table (7).

Unsupervised classification techniques

Unsupervised Classification techniques do not require any training data or prior knowledge, and their main goal

Table 3 Examples of application of clustering methods towards SDGs using EO data

SDGs	Field	Main finding	References
SDG 2 (Zero Hunger)	Agriculture	The proposed methodology based on K-Means and crop images, had a good performance estimating the rice yield	[51]
SDG 7 (Affordable and Clean Energy)	Renewable energy sources	The choice of the clustering technique plays a crucial function in the forecasting of the gross wind power output	[74]
SDG 9 (Industry, Innovation and Infrastructure)	Mining	The results showed that FCM was superior to K-Means and Self-Organising Map for mineral favourability mapping	[90]
SDG 11 (Sustainable Cities and Communities)	Land change	The proposed approach based on K-Means, demonstrated better detection accuracies and visual performance for land cover and land change detection, compared to several methods	[91]
	Seismic	The method analysed was reliable and effective in the identification of sequences of earthquakes	[92]
	Construction	The proposed method used to segment individual buildings had a good performance with datasets acquired from densely built-up areas	[76]
	Land cover	The proposed clustering method outperformed the original approach for remote sensing segmentation in land cover classificatio	[93]
SDG 13 (Climate Action)	Wildfires	The presented algorithm for global burned area mapping was capable to adapt to different ecosystems and spatial resolution data	[54]
	Geomorphology	The proposed DBSCAN methodology for geomorphological analysis allowed the detection of movements of a rock glacier	[94]
	Climate	The techniques used such as K-Means and DBSCAN demonstrated their suitability for predicting climate types	[53]
SDG 14 (Life Below Water)	Sandbars	The proposed algorithm demonstrated a high potential to be used for the extraction of sandbars positions	[95]
SDG 15 (Life on Land)	Soil degradation	Assessment of spatial variability and mapping of soil properties provide an important link in identifying soil degradation spots	[96]
	Agriculture	Optimised kernel-based FCM gave more accurate agriculture crop maps when compared with the classical FCM and K-Means	[97]
SDG 17 (Partnerships for the Goals)	Sustainability level	The results obtained using Hierarchical Cluster Analysis showed that Sweden has the highest level of sustainability among the European countries; while, Greece, Bulgaria and Romania were the countries with the lowest performance	[98]

is to group image pixels or sub-pixels into unlabelled classes [11]. Table 8 lists some recent examples regarding the application of unsupervised classification techniques.

Image segmentation object-based classification

The image segmentation OB classification is used to identify objects based on their proprieties or features. These techniques were developed to emulate the human visual interpretation. Some applications of OB techniques are presented in Table 9.

The success cases presented in Tables 2, 3, 4, 5 and in Sect. "Methodologies and techniques for EO imagery analysis", demonstrate that the contribution of ML is crucial towards the analysis of data provided by EO sources. The synergy between EO and ML can be viewed as an important tool to support a wide variety of SDGs and fields at a global scale and enhance their level of implementation, effectiveness and efficiency. Some of the most common SDGs presented in this paper, which benefits from the synergy EO-ML are: SDG 11, 15 and 9; and the most common fields are Agriculture, Land Cover and Pollution.

Conclusions

Sustainability is an unavoidable aspect for the development of societies and countries; it leads to the development of SDGs and, hence, is crucial to the future of the planet. SDGs are unique as they cover issues that affect

Table 4 Examples of application of regression methods towards SDGs using EO data

SDGs	Field	Main finding	References
SDG 2 (Zero Hunger)	Agriculture	Results increased the potential of using Sentinel-2 to obtain cotton Leaf Area Index and comparison of methods showed that the Gradient Boosting RT was the best	[100]
		Estimate the crop yield, at a pixel level, using ML proved to be an accurate approach	[73]
		The results obtained from the comparison of methods showed that Boosted RT was the best to predict maize yield	[101]
SDG 3 (Good Health and Well-Being)	Spread of diseases	By mapping the relationship between EO variables and vector population, the proposed RF Regression methodology was able to predict the temporal distribution of yellow fever mosquito populations	[102]
SDG 6 (Clean Water and Sanitation)	Water quality	Landsat 7 images are a solid option for assessing water quality characteristics	[55]
SDG 7 (Affordable and Clean Energy)	Renewable energy sources	During Spring and Autumn is harder to predict the hourly solar irradiation compared to Winter and Summer	[103]
SDG 9 (Industry, Innovation and Infrastructure)	Pollution	RT effectively estimates carbon dynamics and allowed the analysis of its impacts on meteorology and vegetation	[49]
		The improved GPR had a high accuracy compared to the original GPR and other methods predicting the CO_2 emissions	[104]
SDG 11 (Sustainable Cities and Communities)	Land cover	RF Regression was very accurate (96%) in delineating house-attached, semi-public and public green spaces	[105]
SDG 13 (Climate Action)	Drought	The use of ML to acquire the *Normalised Microwave Reflection Index* is an effective way to monitor the variation of vegetation water content to predict droughts	[106]
SDG 14 (Life Below Water)	Freshwater habitat	Geographically Weighted Regression technique was accurate in the estimation of stream bathymetry of water with a depth less than 1 m	[107]
SGD 15 (Life on Land)	Terrestrial ecosystem	The best performance, to obtain the latent heat evaporation using a small dataset, was achieved by Kernel Ridge Regression, and using a large dataset, was achieved by Bagging RT	[108]
	Grassland	Vegetation indices acquired from Sentinel 2 have high potential concerning grasslands productivity, management, monitoring and conservation	[109]
	Landslide	Catchment map units and model selection are crucial for the performance of landslide susceptibility maps	[110]

all communities and reaffirm the international commitment to eradicate poverty, hunger and inequalities to build a more sustainable, prosperous and safer planet for all humanity.

This paper highlights the importance of monitoring the SD by means of EO and ML and enhances their fundamental role in pursuing those goals. Monitorisation aspects related to SD, such as poverty, nutrition, health conditions and inequalities have leveraged EO data collection methods. EO is possibly the most cost-effective technology as it is able to provide data at a global level and therefore enabling a global perspective of the SDGs. EO data plays a critical role in promoting equity among developed and developing countries since it grants worldwide data access despite their development level. EO data analysis, which often involves identifying features of interest within large amounts of information (Classification, Clustering, Regression or Dimension Reduction problems), gets even more powerful through the application of ML methods using different methodologies such as Empirical and Semi-Empirical modelling, Sub-PB, PB, Super-PB or even OB techniques.

This extensive review looked at different ML categories to handle EO data to tackle different SDGs. It can be concluded that all ML categories can contribute to a wide variety of SDGs and fields—The Classification category covers the SDGs 2, 6, 8, 11, 13, 14 and 15, and fields such as Agriculture, Land Use and Forests; the Clustering category covers the SDGs 2, 7, 9, 11, 13, 14, 15 and 17, and fields such as Construction, Natural

Table 5 Examples of application of dimension reduction methods towards SDGs using EO data

SDGs	Field	Main finding	References
SGD 2 (Zero Hunger)	Agriculture	Partial Least Square Regression was applied with success, as a FS method, on crop yield estimation	[115]
		The FS results demonstrated that the proposed Maximum Separability and Minimum Dependency method was more accurate than filter methods	[116]
SGD 6 (Clean Water and Sanitation)	Water resources	The proposed approach proved to be effective and accurate to assess water resources at catchment scale	[117]
	Water Sources	Stepwise Discriminant Analysis and PCA improved the accuracy of water source recognition	[118]
SDG 7 (Affordable and Clean Energy)	Electricity	The proposed method improved the forecasting of electricity price and it was more accurate than the Independent Electricity System Operator prediction	[119]
SGD 9 (Industry, Innovation and Infrastructure)	Structural Reliability	The Bivariate Dimension Reduction Method proved to be effective for structural reliability analysis	[120]
SGD 11 (Sustainable Cities and Communities)	Land Cover	The results demonstrated that FS improves the classification accuracy of land cover classification	[121]
		The proposed method demonstrates better results compared to other methods for land cover classification in almost all tests	[122]
		Dimensionality Reduction was considered a key step in the land cover classification process	[123]
		The experiments shown that the impervious surface extraction accuracy of Classification and Regression Tree was higher than Seperability and Thresholds algorithm	[124]
	Land use	FS with Classification Optimisation Score metric reduces the processing time and produces higher classification accuracy for land use and land cover classification using VHR data	[125]
SDG 13 (Climate Action)	Pollution	The new Dimension Reduction method demonstrated to be a powerful approach to optimise the knowledge that emerges from atmospheric observations of N_2O	[126]
SGD 15 (Life on Land)	Forest	Proposed a FS SVM-Recursive Feature Elimination method to explore the relationship between the biomass and parameters derived from Landsat-8 imagery. The results demonstrated that this method was able to accurately estimate the aboveground biomass.	[127]
	Terrestrial ecosystem	FS methods allow the extraction of valuable information to create accurate maps of areas infested by invasive plant species	[128]

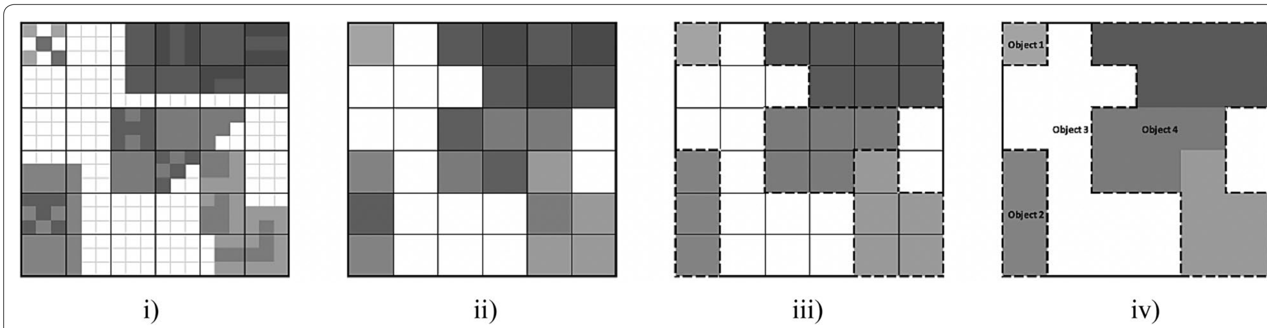

Fig. 4 i) Sub-Pixel-Based; ii) Pixel-Based; iii) Super-Pixel-Based and iv) Object-Based Technique

Table 6 Examples of application of empirical and semi-empirical models towards SDGs using EO data

SDGs	Field	Main finding	References
SDG 2 (Zero Hunger)	Soil condition	Proposed a Semi-Empirical model, which demonstrated its suitability, to reconstruct the signal from Signal-to-Noise Ratio data and simultaneous acquire information that is influenced by soil moisture	[139]
SDG 7 (Affordable and Clean Energy)	Renewable energy sources	Developed a semi-empirical model to forecast the monthly average of solar radiation per hour. The results demonstrated that the estimated value was in agreement with the measurements	[140]
SDG 15 (Life on Land)	Forest	Proposed the use of a Semi-Empirical model with images from RADARSAT-2 to acquire features from the surface of tropical forests. The framework achieved an accuracy of $\cong$83%	[141]
	Invasive plants	Compared supervised and unsupervised image classifiers for mapping a cactus plant, and the results showed that the supervised classifiers were more accurate than the unsupervised classifiers	[142]

Table 7 Examples of application of supervised classification techniques towards SDGs using EO data

SDGs	Field	Main finding	References
SGD 11 (Sustainable Cities and Communities)	Land use	Tested Sub-PB and Super-PB methodologies to map green spaces. The results showed that Super-PB approach was better for dense urban, sub-urban and rural subsets. However, for lower-resolution images, the Sub-PB approach performed better for dense urban and sub-urban subsets	[136]
	Land change	Developed two CNN approaches: Early Fusion and Siamese Network to detect changes in pairs of images. Overall, the results proved that Siamese Network approach was the most accurate	[144]
SGD 12 (Responsible Consumption and Production)	Consumption	Proved that CNNs combined with high-resolution images represent a precise and cost-effective methodology to calculate consumption expenditure and wealth in developing countries	[145]
SDG 15 (Life on Land)	Land cover	Analysed 15 years of research on supervised classification methods and found that SVM was the most accurate among NN, RF and Decision Tree	[146]

Table 8 Examples of application of unsupervised classification techniques towards SDGs using EO data

SDGs	Field	Main finding	Reference
SGD 11 (Sustainable Cities and Communities)	Land cover	Used and compared three methods to classify ground vegetation covers using data acquired by IKONOS satellite. The comparison demonstrated that all methods are very accurate (more than 90% of accuracy); however, the two-step method achieved the best results	[147]
	Land change	Proposed an unsupervised method with an OB approach to improve the detection of changes using high-resolution images. This methodology achieved better results in comparison to other methods	[148]
SGD 15 (Life on Land)	Invasive plants	Developed an unsupervised method to detect and map invasive plants using RFs, which proved to be a successful approach	[149]
	Landslide	Compared an unsupervised PB and OB approach for landslide detection using VHR images and concluded that OB performed better than PB	[150]

Table 9 Examples of application of image segmentation object-based classification towards SDGs using EO data

SDGs	Field	Main finding	References
SGD 11 (Sustainable Cities and Communities)	Land cover	Compared four OB classifiers for the classification of a suburban area with data provided by Landsat-8 and proved that SVM had the best performance among all	[151]
	Land use	Proposed OB approach for urban land use classification using VHR images	[152]
SGD 15 (Life on Land)	Land cover	Tested the performance of PB and OB classification with a hyperspectral dataset and found that OB was better than PB approach	[153]
	Land use	Compared an OB and PB approach using aero photogrammetric images and the results showed that OB classifier performed better compared to PB	[154]

Disasters and Renewable Energy; the Regression category covers the SDGs 2, 3, 6, 7, 9, 11, 13, 14 and 15, and the fields Water Quality, Pollution and Freshwater; and the Dimension Reduction category covers the SDGs 3, 6, 7, 9, 11, 13 and 15, and the fields Land Cover, Electricity and Software.

Thus, the overall findings confirm the significance of EO and ML in pursuing the goals of SD providing an overview of methods and techniques that sustain the achievement of SDGs. Lastly, the applicability and efficiency of specific ML methods used to analyse EO data, such as Random Forest (RF), Support Vector Machine (SVM) and Neural Network (NN), should be further explored to sustain a more consensual and reliable development/improvement of tools to support SDGs.

Abbreviations

BR: Brundtland Report; CNN: Convolutional Neural Network; DBSCAN: Density-Based Spatial Clustering of Applications with Noise; EO: Earth observation; FCM: Fuzzy C-means; FS: Feature selection; GPR: Gaussian process regression; K-NN: K-Nearest neighbour; LLE: Locally linear embedding; MDG: Millennium Development Goal; ML: Machine learning; NN: Neural network; OB: Object based; PB: Pixel based; PCA: Principal component analysis; RF: Random forest; RT: Regression tree; SD: Sustainable development; SDG: Sustainable Development Goal; SVM: Support vector machine; t-SNE: t-Distributed stochastic neighbour embedding; VHR: Very high resolution; WCED: World Commission on Environment and Development.

Acknowledgements

This work was partially supported by Fundação para a Ciência e a Tecnologia, Portugal, Project UIDB/04005/2020 and Project CMUP-ERI/TIC/0045/2014.

Authors' contributions

BF: conceptualisation, methodology, investigation, writing—original draft, writing—review and editing; MI and RS: writing—review and editing; supervision. All authors read and approved the final manuscript.

Funding

This work was supported through Project +Atlantic CMUP-ERI/TIC/0045/2014 and Project UIDB/04005/2020.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] ISQ-Intelligent and Digital Systems R&Di, Instituto de Soldadura e Qualidade, Rua do Mirante, 258, 4415-491 Grijó, Portugal. [2] ISQ-Low Carbon and Resource Efficiency, R&Di, Instituto de Soldadura e Qualidade, Rua do Mirante, 258, 4415-491 Grijó, Portugal. [3] Universidade Lusíada–Norte, 4760-108 Vila Nova de Famalicão, Portugal. [4] COMEGI-Centro de Investigação em Organizações, Mercados e Gestão Industrial, Lisbon, Portugal.

Received: 2 June 2020 Accepted: 3 September 2020
Published online: 21 September 2020

References

1. Meadows DH, Meadows D, Randers J, Behrens WW III (1972) The limits to growth: a report to the club of Rome. Formaliz Math. https://doi.org/10.2478/v10037-009-0024-8
2. Mesarovic M, Pestel E (1975) Mankind at the turning point: the second report of the club of Rome. Econ Anal Policy. https://doi.org/10.1016/S0313-5926(76)50005-1
3. WCED (1987) Report of the World Commission on Environment and Development: Our Common Future: Report of the World Commission on Environment and Development
4. United Nations (2000) United Nations Millennium Declaration. Gen Assem 9
5. United Nations (2015) Transforming our world: the 2030 agenda for sustainable development. N Era Glob Heal. https://doi.org/10.1891/9780826190123.ap02
6. United Nations (2015) The Millennium Development Goals Report 2015. New York. https://doi.org/10.18356/6cd11401-en
7. Moucheraud C, Owen H, Singh NS et al (2016) Countdown to 2015 country case studies: what have we learned about processes and progress towards MDGs 4 and 5? BMC Public Health 16:794. https://doi.org/10.1186/s12889-016-3401-6
8. United Nations Development Programme (2017) Background of the sustainable development goals. In: United Nations Dev. Program. https://www.undp.org/content/undp/en/home/sustainable-development-goals/background.html. Accessed 15 Feb 2019
9. Gusmão Caiado RG, Leal Filho W, Quelhas OLG et al (2018) A literature-based review on potentials and constraints in the implementation of the sustainable development goals. J Clean Prod. 198:1276–1288
10. GEO (2017) Earth Observations in support of the 2030 Agenda for Sustainable Development. Retrieved from https://www.earthobservations.org/documents/publications/201703_geo_eo_for_2030_agenda.pdf
11. Andries A, Morse S, Murphy R et al (2019) Translation of Earth observation data into sustainable development indicators: an analytical framework. Sustain Dev 27:366–376. https://doi.org/10.1002/sd.1908
12. United Nations (2017) Earth observations for official statistics: satellite imagery and geospatial data task team report. Retrieved from https://unstats.un.org/bigdata/taskteams/satellite/UNGWG_Satellite_Task_Team_Report_WhiteCover.pdf
13. Xie M, Jean N, Burke M et al (2016) Testing the race model inequality in redundant stimuli with variable onset asynchrony. J Exp Psychol Hum Percept Perform 35:575–579. https://doi.org/10.1037/a0013620
14. Murthy K, Shearn M, Smiley BD, et al (2014) SkySat-1: very high-resolution imagery from a small satellite. In: Meynart R, Neeck SP, Shimoda H (eds). International Society for Optics and Photonics, Bellingham
15. Landry T, Sotir M, Rajotte J-F, et al (2019) Applying machine learning to earth observations in a standards based workflow. IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp 5567–5570
16. Turner GM (2008) A comparison of The Limits to Growth with 30 years of reality. Glob Environ Chang 18:397–411. https://doi.org/10.1016/j.gloenvcha.2008.05.001
17. Coscieme L, Sutton P, Mortensen LF et al (2019) Overcoming the myths of mainstream economics to enable a new wellbeing economy. Sustainability 11:4374. https://doi.org/10.3390/su11164374
18. Meadows DH, Randers J, Meadows DL (2004) Limits to growth: the 30-year update. Chelsea Green Publishing, Hartford
19. Meadows DL, Meadows D, Randers J (1992) Beyond the limits: confronting global collapse. Chelsea Green Publishing, Envisioning a Sustainable Future
20. Bermejo R, Arto I, Hoyos D (2010) Sustainable development in the brundtlant report and its distortion : implications for development economics and international cooperation
21. Klarin T (2018) The concept of sustainable development: from its beginning to the contemporary issues. Zagreb Int Rev Econ Bus 21:67–94. https://doi.org/10.2478/zireb-2018-0005
22. United Nations (1969) A study of the capacity of the United Nations Development Systems. Retrieved from http://digitallibrary.un.org/record/695861

23. European Environment Agency (1997) EEA Annual Report. Copenhagen. Retrieved from https://www.eea.europa.eu/publications/eea-annual-report-1997
24. United Nations (2016) Millennium Development Goals. https://www.un.org/millenniumgoals/. Accessed 17 Feb 2020
25. Rey J, Sachs D (2012) For more on the report by the High-level Panel on Global From Millennium Development Goals to Sustainable Development Goals
26. Salvia AL, Leal Filho W, Brandli LL, Griebeler JS (2019) Assessing research trends related to Sustainable Development Goals: local and global issues. J Clean Prod 208:841–849. https://doi.org/10.1016/j.jclepro.2018.09.242
27. United Nations (2019) Sustainable development knowledge platform. https://sustainabledevelopment.un.org/. Accessed 17 Feb 2020
28. Coscieme L, Mortensen LF, Anderson S et al (2020) Going beyond Gross Domestic Product as an indicator to bring coherence to the Sustainable Development Goals. J Clean Prod. 248:119232
29. United Nations (2015) Adoption of the Paris Agreement. Framework Convention on Climate Change. Retrieved from https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf
30. Anderson K, Ryan B, Sonntag W et al (2017) Earth observation in service of the 2030 Agenda for sustainable development. Geo-spatial Inf Sci 20:77–96. https://doi.org/10.1080/10095020.2017.1333230
31. European Space Agency (ESA) (2018) Satellite Earth Observations in support of the Sustainable Development Goals
32. Kim BY, Lee KT (2018) Radiation component calculation and energy budget analysis for the Korean Peninsula Region. Remote Sens 10:1147. https://doi.org/10.3390/rs10071147
33. ESA (2018) Space Debris By The Numbers. In: ESA Website. https://www.esa.int/Our_Activities/Operations/Space_Debris. Accessed 4 May 2020
34. NASA (2019) EarthData. https://earthdata.nasa.gov/. Accessed 25 Oct 2019
35. Yang C, Yu M, Li Y et al (2019) Big Earth data analytics: a survey. Big Earth Data 3:83–107. https://doi.org/10.1080/20964471.2019.1611175
36. Faghmous JH, Kumar V (2014) A Big Data guide to understanding climate change: the case for theory-guided data science. Big Data 2:155–163. https://doi.org/10.1089/big.2014.0026
37. Jeltsch F, Bonte D, Pe'er G et al (2013) Integrating movement ecology with biodiversity research—exploring new avenues to address spatiotemporal biodiversity dynamics. Mov Ecol 1:1–13. https://doi.org/10.1186/2051-3933-1-6
38. Holloway J, Mengersen K, Helmstedt K (2018) Spatial and machine learning methods of satellite imagery analysis for Sustainable Development Goals
39. Yu B, Liu H, Wu J et al (2010) Automated derivation of urban building density information using airborne LiDAR data and object-based method. Landsc Urban Plan 98:210–219. https://doi.org/10.1016/j.landurbplan.2010.08.004
40. Ottinger M, Clauss K, Kuenzer C (2018) Opportunities and challenges for the estimation of aquaculture production based on earth observation data. Remote Sens 10:1–24. https://doi.org/10.3390/rs10071076
41. Fabre S, Briottet X, Lesaignoux A (2015) Estimation of soil moisture content from the spectral reflectance of bare soils in the 0.4–2.5 µm domain. Sensors 15:3262–3281. https://doi.org/10.3390/s150203262
42. Govender M, Chetty K, Bulcock H (2007) A review of hyperspectral remote sensing and its application in vegetation and water resource studies. Water Res Comm 33:145–152
43. Zhang G, Strøm JS, Blanke M, Braithwaite I (2006) Spectral signatures of surface materials in pig buildings. Biosyst Eng 94:495–504. https://doi.org/10.1016/j.biosystemseng.2006.05.002
44. Ose K, Corpetti T, Demagistri L (2016) Multispectral satellite image processing. Optical remote sensing of land surface. Elsevier, pp 57–124
45. Food and Agriculture Organization (2016) The State of Food and Agriculture: Climate change, agriculture and food security
46. GEO (2019) Earth Observations and geospatial information: supporting official statistics in monitoring and achieving the 2030 agenda
47. García L, Rodríguez D, Wijnen M, Pakulski I (2016) Earth observation for water resources management: current use and future opportunities for the water sector. World Bank, Washington, DC
48. Sharma R, Ghosh A, Joshi PK (2013) Spatio-temporal footprints of urbanisation in Surat, the Diamond City of India (1990–2009). Environ Monit Assess 185:3313–3325. https://doi.org/10.1007/s10661-012-2792-9
49. Boyte SP, Wylie BK, Howard DM et al (2018) Estimating carbon and showing impacts of drought using satellite data in regression-tree models. Int J Remote Sens 39:374–398. https://doi.org/10.1080/01431161.2017.1384592
50. Poursanidis D, Topouzelis K, Chrysoulakis N (2018) Mapping coastal marine habitats and delineating the deep limits of the Neptune's seagrass meadows using very high resolution Earth observation data. Int J Remote Sens 39:8670–8687. https://doi.org/10.1080/01431161.2018.1490974
51. Reza MN, Na IS, Baek SW, Lee KH (2019) Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images. Biosyst Eng 177:109–121. https://doi.org/10.1016/j.biosystemseng.2018.09.014
52. Puletti N, Chianucci F, Castaldi C (2018) Use of Sentinel-2 for forest classification in Mediterranean environments. Ann Silvic Res 42:32–38. https://doi.org/10.12899/ASR-1463
53. Sathiaraj D, Huang X, Chen J (2019) Predicting climate types for the Continental United States using unsupervised clustering techniques. Environmetrics. Wiley, New Jersey
54. Lizundia-Loiola J, Otón G, Ramo R, Chuvieco E (2020) A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data. Remote Sens Environ 236:111493. https://doi.org/10.1016/j.rse.2019.111493
55. Sharma B, Kumar M, Denis DM, Singh SK (2019) Appraisal of river water quality using open-access earth observation data set: a study of river Ganga at Allahabad (India). Sustain Water Resour Manag 5:755–765. https://doi.org/10.1007/s40899-018-0251-7
56. Firozjaei MK, Sedighi A, Argany M et al (2019) A geographical direction-based approach for capturing the local variation of urban expansion in the application of CA-Markov model. Cities 93:120–135. https://doi.org/10.1016/j.cities.2019.05.001
57. Wang L, Dong Q, Yang L et al (2019) Crop classification based on a novel feature filtering and enhancement method. Remote Sens 11:455. https://doi.org/10.3390/rs11040455
58. Foody G, Ling F, Boyd D et al (2019) Earth observation and machine learning to meet sustainable development goal 8.7: mapping sites associated with slavery from space. Remote Sens 11:266. https://doi.org/10.3390/rs11030266
59. Global Working Group on Big Data (2017) Satellite imagery and geo-spatial dataglobal working group on big data. (2017). Satellite imagery and geo-spatial data. https://unstats.un.org/bigdata/taskteams/satellite/
60. Lary DJ, Zewdie GK, Liu X et al (2018) Machine learning applications for earth observation. Earth observation open science and innovation. Springer, Cham, pp 165–218. https://doi.org/10.1007/978-3-319-65633-5
61. Transon J, d'Andrimont R, Maugnard A, Defourny P (2018) Survey of hyperspectral Earth Observation applications from space in the Sentinel-2 context. Remote Sens 10:157. https://doi.org/10.3390/rs10020157
62. Abelson B, Varshney KR, Sun J (2014) Targeting direct cash transfers to the extremely poor. 1563–1572. https://doi.org/10.1145/2623330.2623335
63. Varshney KR, Chen GH, Abelson B et al (2015) Targeting villages for rural development using satellite image analysis. Big Data 3:41–53. https://doi.org/10.1089/big.2014.0061
64. Berry M, Mohamed A, Yap BW (2019) Supervised and unsupervised learning for data science. Springer, Cham
65. Baştanlar Y, Özuysal M (2014) Introduction to machine learning. Methods Mol Biol 1107:105–128. https://doi.org/10.1007/978-1-62703-748-8_7
66. Yasodha P, Kannan M (2011) Analysis of a population of diabetic patients databases in WEKA tool. Int J Sci Eng Res. 2(5):1–5
67. Osband I, Doron Y, Hessel M, et al (2020) Behaviour suite for reinforcement learning. int conf learn represent
68. Zhu H, Yu J, Gupta A, et al (2020) The ingredients of real-world robotic reinforcement learning. Int Conf Learn Represent
69. Samuel A (1959) Some studies in machine learning using the game of checkers. IBM J 3:210–229. https://doi.org/10.1147/rd.33.0210

70. Ongsulee P (2018) Artificial intelligence, machine learning and deep learning. In: International Conference on ICT and Knowledge Engineering. IEEE Computer Society, pp 1–6
71. Jakhar D, Kaur I (2020) Artificial intelligence, machine learning and deep learning: definitions and differences. Clin Exp Dermatol 45:131–132. https://doi.org/10.1111/ced.14029
72. Luxton DD (2016) An introduction to artificial intelligence in behavioral and mental health care. Artificial intelligence in behavioral and mental health care. Elsevier, New Jersey, pp 1–26
73. Kamir E, Waldner F, Hochman Z (2020) Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. ISPRS J Photogramm Remote Sens 160:124–135. https://doi.org/10.1016/j.isprsjprs.2019.11.008
74. Adedeji PA, Akinlabi S, Madushele N, Olatunji OO (2020) Wind turbine power output very short-term forecast: a comparative study of data clustering techniques in a PSO-ANFIS model. J Clean Prod 254:1–16. https://doi.org/10.1016/j.jclepro.2020.120135
75. Zhang G, Ghamisi P, Zhu XX (2019) Fusion of heterogeneous earth observation data for the classification of local climate zones. IEEE Trans Geosci Remote Sens 57:7623–7642. https://doi.org/10.1109/TGRS.2019.2914967
76. Huang X, Cao R, Cao Y (2019) A density-based clustering method for the segmentation of individual buildings from filtered airborne LiDAR point clouds. J Indian Soc Remote Sens 47:907–921. https://doi.org/10.1007/s12524-018-0911-y
77. Damgacioglu H, Celik E, Celik N (2019) Estimating gene expression from high-dimensional DNA methylation levels in cancer data: a bimodal unsupervised dimension reduction algorithm. Comput Ind Eng 130:348–357. https://doi.org/10.1016/j.cie.2019.02.038
78. Holloway J, Mengersen K (2018) Statistical machine learning methods and remote sensing for sustainable development goals: a review. Remote Sens 10(9):1365
79. Ferreira B, Silva RG, Pereira V (2017) Feature selection using non-binary decision trees applied to condition monitoring. IEEE Int Conf Emerg Technol Fact Autom ETFA Cyprus, Limassol 1:1–7. https://doi.org/10.1109/ETFA.2017.8247642
80. Chandrashekar G, Sahin F (2014) A survey on feature selection methods. Comput Electr Eng 40:16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024
81. Vuolo F, Neuwirth M, Immitzer M et al (2018) How much does multi-temporal Sentinel-2 data improve crop type classification? Int J Appl Earth Obs Geoinf 72:122–130. https://doi.org/10.1016/j.jag.2018.06.007
82. Zhong L, Hu L, Zhou H (2019) Deep learning based multi-temporal crop classification. Remote Sens Environ 221:430–443. https://doi.org/10.1016/j.rse.2018.11.032
83. Zhang M, Chen F, Tian B, Liang D (2019) Multi-temporal SAR image classification of coastal plain wetlands using a new feature selection method and random forests. Remote Sens Lett 10:312–321. https://doi.org/10.1080/2150704X.2018.1528397
84. Helber P, Bischke B, Dengel A, Borth D (2019) Eurosat: a novel dataset and deep learning benchmark for land use and land cover classification. IEEE J Sel Top Appl Earth Obs Remote Sens 12:2217–2226. https://doi.org/10.1109/JSTARS.2019.2918242
85. Kuffer M, Wang J, Nagenborg M et al (2018) The scope of earth-observation to improve the consistency of the SDG slum indicator. ISPRS Int J Geo-Inform 7:428. https://doi.org/10.3390/ijgi7110428
86. Schäfer P, Pflugmacher D, Hostert P, Leser U (2018) Classifying land cover from satellite images using time series analytics. CEUR Workshop Proc 2083:10–15
87. Ahmed AM, Ibrahim SK, Yacout S (2019) Hyperspectral image classification based on logical analysis of data. IEEE Aerosp Conf Proc. https://doi.org/10.1109/AERO.2019.8742023
88. dos Reis AA, Carvalho MC, de Mello JM et al (2018) Spatial prediction of basal area and volume in Eucalyptus stands using Landsat TM data: an assessment of prediction methods. New Zeal J For Sci 48:1. https://doi.org/10.1186/s40490-017-0108-0
89. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction. J Am Stat Assoc 99:567–567. https://doi.org/10.1198/jasa.2004.s339
90. Rezapour MJ, Abedi M, Bahroudi A, Rahimi H (2019) A clustering approach for mineral potential mapping: a deposit-scale porphyry copper exploration targeting. Geopersia. https://doi.org/10.22059/GEOPE.2019.284414.648488
91. Lv Z, Liu T, Shi C et al (2019) Novel land cover change detection method based on k-means clustering and adaptive majority voting using bitemporal remote sensing images. IEEE Access 7:34425–34437. https://doi.org/10.1109/ACCESS.2019.2892648
92. Peresan A, Gentili S (2018) Seismic clusters analysis in Northeastern Italy by the nearest-neighbor approach. Phys Earth Planet Inter 274:87–104. https://doi.org/10.1016/j.pepi.2017.11.007
93. Chen S, Sun T, Yang F et al (2018) An improved optimum-path forest clustering algorithm for remote sensing image segmentation. Comput Geosci 112:38–46. https://doi.org/10.1016/j.cageo.2017.12.003
94. Micheletti N, Tonini M, Lane SN (2017) Geomorphological activity at a rock glacier front detected with a 3D density-based clustering algorithm. Geomorphology 278:287–297. https://doi.org/10.1016/j.geomorph.2016.11.016
95. Tatui F, Constantin S (2019) Nearshore sandbar crest position dynamics analysed based on Earth Observation data. Remote Sens Environ Rev. https://doi.org/10.1016/j.rse.2019.111555
96. Verma RR, Manjunath BL, Singh NP et al (2018) Soil mapping and delineation of management zones in the Western Ghats of coastal India. L Degrad Dev 29:4313–4322. https://doi.org/10.1002/ldr.3183
97. Tamiminia H, Homayouni S, McNairn H, Safari A (2017) A particle swarm optimized kernel-based clustering method for crop mapping from multi-temporal polarimetric L-band SAR observations. Int J Appl Earth Obs Geoinf 58:201–212. https://doi.org/10.1016/j.jag.2017.02.010
98. Drastichová M, Filzmoser P (2019) Assessment of sustainable development using cluster analysis and principal component analysis. Probl Ekorozwoju 14:7–24
99. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313:504–507. https://doi.org/10.1126/science.1127647
100. Mao H, Meng J, Ji F et al (2019) Comparison of machine learning regression algorithms for cotton leaf area index retrieval using Sentinel-2 spectral bands. Appl Sci. https://doi.org/10.3390/app9071459
101. Aghighi H, Azadbakht M, Ashourloo D et al (2018) Machine learning regression techniques for the silage maize yield prediction using time-series images of landsat 8 OLI. IEEE J Sel Top Appl Earth Obs Remote Sens 11:4563–4577. https://doi.org/10.1109/JSTARS.2018.2823361
102. Mudele O, Bayer FM, Zanandrez LFR et al (2020) Modeling the temporal population distribution of ae. mosquito using big earth observation data. IEEE Access 8:14182–14194. https://doi.org/10.1109/ACCESS.2020.2966080
103. Benali L, Notton G, Fouilloy A et al (2019) Solar radiation forecasting using artificial neural network and random forest methods: application to normal beam, horizontal diffuse and global components. Renew Energy 132:871–884. https://doi.org/10.1016/j.renene.2018.08.044
104. Fang D, Zhang X, Yu Q et al (2018) A novel method for carbon dioxide emission forecasting based on improved Gaussian processes regression. J Clean Prod 173:143–150. https://doi.org/10.1016/j.jclepro.2017.05.102
105. Haase D, Jänicke C, Wellmann T (2019) Front and back yard green analysis with subpixel vegetation fractions from earth observation data in a city. Landsc Urban Plan 182:44–54. https://doi.org/10.1016/j.landurbplan.2018.10.010
106. Yuan Q, Li S, Yue L et al (2019) Monitoring the variation of vegetation water content with machine learning methods: point-surface fusion of MODIS products and GNSS-IR observations. Remote Sens 11:1440. https://doi.org/10.3390/rs11121440
107. Kim JS, Baek D, Seo IW, Shin J (2019) Retrieving shallow stream bathymetry from UAV-assisted RGB imagery using a geospatial regression method. Geomorphology 341:102–114. https://doi.org/10.1016/j.geomorph.2019.05.016
108. Carter C, Liang S (2019) Evaluation of ten machine learning methods for estimating terrestrial evapotranspiration from remote sensing. Int J Appl Earth Obs Geoinf 78:86–92. https://doi.org/10.1016/j.jag.2019.01.020
109. Guerini Filho M, Kuplich TM, De Quadros FLF (2020) Estimating natural grassland biomass by vegetation indices using Sentinel 2 remote sensing data. Int J Remote Sens 41:2861–2876. https://doi.org/10.1080/01431161.2019.1697004

110. Chu L, Wang L-J, Jiang J et al (2019) Comparison of landslide susceptibility maps using random forest and multivariate adaptive regression spline models in combination with catchment map units. Geosci J 23:341–355. https://doi.org/10.1007/s12303-018-0038-8
111. Bharti KK, Singh PK (2015) Hybrid dimension reduction by integrating feature selection with feature extraction method for text clustering. Expert Syst Appl 42:3105–3114. https://doi.org/10.1016/j.eswa.2014.11.038
112. Hira ZM, Gillies DF (2015) A review of feature selection and feature extraction methods applied on microarray data. Adv Bioinform. https://doi.org/10.1155/2015/198363
113. Silva RG, Wilcox SJ (2019) Feature evaluation and selection for condition monitoring using a self-organizing map and spatial statistics. Artif Intell Eng Des Anal Manuf 33:1–10. https://doi.org/10.1017/S0890060417000518
114. Van Der Maaten LJP, Postma EO, Van Den Herik HJ (2009) Dimensionality reduction: a comparative review. J Mach Learn Res 10:1–41. https://doi.org/10.1080/13506280444000102
115. Fletcher S, Alemohammad H, Figueroa AJ, Entekhabi D (2019) Characterizing farm-scale variability in maize yields in West Africa by integrating optical and passive microwave earth observation data with a process model. In: AGU Fall Meeting. p GC31C-02
116. Khosravi I, Safari A, Homayouni S (2018) MSMD: maximum separability and minimum dependency feature selection for cropland classification from optical and radar data. Int J Remote Sens 39:2159–2176. https://doi.org/10.1080/01431161.2018.1425564
117. Owen NE, Liuzzo L (2019) Impact of land use on water resources via a Gaussian process emulator with dimension reduction. J Hydroinformatics 21:411–426. https://doi.org/10.2166/hydro.2019.067
118. Hou E, Wen Q, Che X et al (2020) Study on recognition of mine water sources based on statistical analysis. Arab J Geosci. https://doi.org/10.1007/s12517-019-4984-x
119. Jahangir H, Tayarani H, Baghali S et al (2019) A novel electricity price forecasting approach based on dimension reduction strategy and rough artificial neural networks. IEEE Trans Ind Informatics. https://doi.org/10.1109/tii.2019.2933009
120. Xu J, Dang C (2019) A new bivariate dimension reduction method for efficient structural reliability analysis. Mech Syst Signal Process 115:281–300. https://doi.org/10.1016/j.ymssp.2018.05.046
121. Dogan T, Uysal AK (2018) The impact of feature selection on urban land cover classification. Int J Intell Syst Appl Eng 6:59–64. https://doi.org/10.18201/ijisae.2018637933
122. Bui QT, Van Pham M, Nguyen QH et al (2019) Whale optimization algorithm and adaptive neuro-fuzzy inference system: a hybrid method for feature selection and land pattern classification. Int J Remote Sens 40:5078–5093. https://doi.org/10.1080/01431161.2019.1578000
123. Stromann O, Nascetti A, Yousif O, Ban Y (2020) Dimensionality Reduction and Feature Selection for Object-Based Land Cover Classification based on Sentinel-1 and Sentinel-2 Time Series Using Google Earth Engine. Remote Sens 12(1):76. https://doi.org/10.3390/RS12010076
124. Fu H, Shao Z, Tu C, Zhang Q (2016) Impacts of feature selection for urban impervious surface extraction using optical image and SAR data. In: Proceedings of 4th International Workshop on Earth Observation and Remote Sensing Applications, EORSA 2016. Institute of Electrical and Electronics Engineers Inc., pp 419–423
125. Georganos S, Grippa T, Vanhuysse S et al (2018) Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. GIScience Remote Sens 55:221–242. https://doi.org/10.1080/15481603.2017.1408892
126. Wells KC, Millet DB, Bousserez N et al (2018) Top-down constraints on global N 2 O emissions at optimal resolution: application of a new dimension reduction technique. Atmos Chem Phys 18:735–756. https://doi.org/10.5194/acp-18-735-2018
127. Qiu A, Yang Y, Wang D, et al (2019) Exploring parameter selection for carbon monitoring based on Landsat-8 imagery of the aboveground forest biomass on Mount Tai. Eur J Remote Sens 53(sup1):4–15. https://doi.org/10.1080/22797254.2019.1686717
128. Kiala Z, Mutanga O, Odindi J, Peerbhay K (2019) Feature selection on sentinel-2 multispectral imagery for mapping a landscape infested by parthenium weed. Remote Sens 11:1892. https://doi.org/10.3390/rs11161892
129. Hussain M, Chen D, Cheng A et al (2013) Change detection from remotely sensed images: from pixel-based to object-based approaches. ISPRS J Photogramm Remote Sens 80:91–106. https://doi.org/10.1016/j.isprsjprs.2013.03.006
130. Lu T, Li S, Fang L et al (2017) From subpixel to superpixel: a novel fusion framework for hyperspectral image classification. IEEE Trans Geosci Remote Sens 55:4398–4411. https://doi.org/10.1109/TGRS.2017.2691906
131. Liu Q, Trinder J, Turner I (2016) A Comparison of sub-pixel mapping methods for coastal areas. ISPRS Ann Photogramm Remote Sens Spat Inf Sci. https://doi.org/10.5194/isprs-annals-III-7-67-2016
132. Mertens K (2008) Towards sub-pixel mapping: design and comparison of techniques
133. Belgiu M, Csillik O (2018) Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sens Environ 204:509–523. https://doi.org/10.1016/j.rse.2017.10.005
134. Ouyang ZT, Zhang MQ, Xie X et al (2011) A comparison of pixel-based and object-oriented approaches to VHR imagery for mapping saltmarsh plants. Ecol Inform 6:136–146. https://doi.org/10.1016/j.ecoinf.2011.01.002
135. Mehra J, Neeru N (2016) A brief review: super-pixel based image segmentation methods. Int J Res Eng 03:8–12
136. Yin W, Yang J (2017) Sub-pixel vs. super-pixel-based greenspace mapping along the urban–rural gradient using high spatial resolution Gaofen-2 satellite imagery: a case study of Haidian District, Beijing, China. Int J Remote Sens 38:6386–6406. https://doi.org/10.1080/01431161.2017.1354266
137. Blaschke T (2010) Object based image analysis for remote sensing. ISPRS J Photogramm Remote Sens 65:2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004
138. Hay G, Castilla G (2006) Object-based image analysis: strengths, weaknesses, opportunities and threats (SWOT). Earth Sci 454:
139. Han M, Zhu Y, Yang D et al (2018) A semi-empirical SNR model for soil moisture retrieval using GNSS SNR data. Remote Sens 10:1–19. https://doi.org/10.3390/rs10020280
140. Charuchittipan D, Choosri P, Janjai S et al (2018) A semi-empirical model for estimating diffuse solar near infrared radiation in Thailand using ground- and satellite-based data for mapping applications. Renew Energy 117:175–183. https://doi.org/10.1016/j.renene.2017.10.045
141. Constantino-Recillas DE, Monsiváis-Huertero A, Jiménez-Escalona JC et al (2018) A semi-empirical model to estimate biophysical parameters in southern Mexico. Int Geosci Remote Sens Symp. https://doi.org/10.1109/IGARSS.2018.8518991
142. Mafanya M, Tsele P, Botai J et al (2017) Evaluating pixel and object based image classification techniques for mapping plant invasions from UAV derived aerial imagery: harrisia pomanensis as a case study. ISPRS J Photogramm Remote Sens 129:1–11. https://doi.org/10.1016/j.isprsjprs.2017.04.009
143. Richards JA (2013) Remote sensing digital image analysis: An introduction, 5th edn. Springer, Berlin, Heidelberg
144. Daudt RC, Le Saux B, Boulch A, Gousseau Y (2018) Urban change detection for multispectral earth observation using convolutional neural networks. Int Geosci Remote Sens Symp. https://doi.org/10.1109/IGARSS.2018.8518015
145. Jean N, Burke M, Xie M et al (2016) Combining satellite imagery and machine learning to predict poverty. Science 353:790–794. https://doi.org/10.1126/science.aaf7894
146. Khatami R, Mountrakis G, Stehman SV (2016) A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. Remote Sens Environ 177:89–100. https://doi.org/10.1016/j.rse.2016.02.028
147. Anchang JY, Ananga EO, Pu R (2016) An efficient unsupervised index based approach for mapping urban vegetation from IKONOS imagery. Int J Appl Earth Obs Geoinf 50:211–220. https://doi.org/10.1016/j.jag.2016.04.001
148. Lv P, Zhong Y, Zhao J, Zhang L (2018) Unsupervised change detection based on hybrid conditional random field model for high spatial

resolution remote sensing imagery. IEEE Trans Geosci Remote Sens 56:4002–4015. https://doi.org/10.1109/TGRS.2018.2819367
149. Peerbhay K, Mutanga O, Lottering R, Ismail R (2016) Mapping Solanum mauritianum plant invasions using WorldView-2 imagery and unsupervised random forests. Remote Sens Environ 182:39–48. https://doi.org/10.1016/j.rse.2016.04.025
150. Keyport RN, Oommen T, Martha TR et al (2018) A comparative analysis of pixel- and object-based detection of landslides from very high-resolution images. Int J Appl Earth Obs Geoinf 64:1–11. https://doi.org/10.1016/j.jag.2017.08.015
151. Shang M, Wang S, Zhou Y et al (2019) Object-based image analysis of suburban landscapes using Landsat-8 imagery. Int J Digit Earth 12:720–736. https://doi.org/10.1080/17538947.2018.1474959
152. Zhang C, Sargent I, Pan X et al (2018) An object-based convolutional neural network (OCNN) for urban land use classification. Remote Sens Environ 216:57–70. https://doi.org/10.1016/j.rse.2018.06.034
153. Kavzoglu T, Tonbul H, Yildiz Erdemir M, Colkesen I (2018) Dimensionality reduction and classification of hyperspectral images using object-based image analysis. J Indian Soc Remote Sens 46:1297–1306. https://doi.org/10.1007/s12524-018-0803-1
154. Piazza GA, Vibrans AC, Liesenberg V, Refosco JC (2016) Object-oriented and pixel-based classification approaches to classify tropical successional stages using airborne high-spatial resolution images. GIScience Remote Sens 53:206–226. https://doi.org/10.1080/15481603.2015.1130589

Publisher's Note